Universidade do Minho
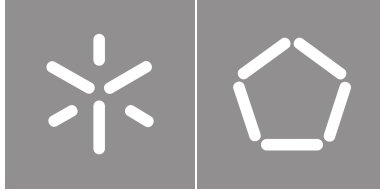Escola de Engenharia

Ana Margarida da Rocha Ferreira

Breast Tumor Segmentation and Classification using Deep Learning Methods

January, 2023

Universidade do Minho
Escola de Engenharia

Ana Margarida da Rocha Ferreira

**Breast Tumor Segmentation and
Classification using Deep Learning Methods**

Masters Dissertation
Master's in Informatics Engineering

Dissertation supervised by:
**Prof. Dr. Paulo Jorge Freitas de Oliveira Novais
Prof. Dr. João Luís Araújo Martins Vilaça**

January, 2023

COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositoriUM of Universidade do Minho.

Acknowledgements

This master thesis is the result of one year of work to which many people contributed. Some had a direct impact, while many others influenced me in a more indirect way.

First and foremost, I would like to thank my supervisor Professor Doctor Paulo Novais, and co-supervisor Professor Doctor João Vilaça, for their scientific orientation and motivation throughout the project. This work is part of the OncoNavigator project, held in the 2Ai Laboratory of the Polytechnic Institute of Cávado and Ave in partnership with the University of Minho, in which I'm grateful to be integrated through a research grant funded by the FCT. This project addresses a fascinating and relevant topic to me, which was crucial for the development of my thesis.

To Pedro Morais, Helena Torres, Bruno Oliveira, João Fonseca, and Augusto Righetti, a major thanks for the availability, knowledge, advice, and continuous support they provided me during the research and development of this project. Thank you also for the opportunity to be part of the 2Ai fantastic team that has allowed me to acquire several skills, experience unique moments, and grow as a person.

To my teammates and laboratory partners, notably Andreia, António, Simão, Raul, Diana, Octávio, Thomas, Bruno, Zé, Luís, Nuno Rodrigues, and Nuno Costa, thanks for the shared moments on those four walls, for the knowledge, team spirit, and friendship. Working with you has been a very enriching experience professionally and personally.

To my friends Sara, Marcelo, Anita, Tânia, Pedro, and Xana, thank you for helping me recharge my batteries to work and for always being here.

Last but not least, I would like to thank my family, specially my parents, my brother, and my sister, for all their unconditional support, motivation, advice, patience, and kindness throughout this journey. Thank you for everything.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Universidade do Minho, Braga, January 2023

_____
(Ana Margarida da Rocha Ferreira)

Resumo

Segmentação e Classificação de Tumores da Mama através de Métodos de Deep Learning

O cancro da mama é o tipo de cancro mais comum e a principal causa de morte por cancro nas mulheres. A deteção atempada é crucial para o sucesso do tratamento e para a redução da taxa de mortalidade. Embora várias modalidades de imagem sejam utilizadas para detetar lesões mamárias, a ultrassonografia tornou-se uma das mais frequentes, uma vez que é segura, portátil, de baixo custo e permite uma examinação em tempo real. Deste modo, a segmentação e classificação automáticas de tumores em imagens de ultrassom da mama podem auxiliar no seu diagnóstico, proporcionando uma segunda opinião aos especialistas. No entanto, realizar uma segmentação e classificação de lesões com precisão, através desta modalidade de imagem, é desafiante devido à má qualidade de imagem e elevada variabilidade das lesões. Recentemente, algoritmos de *deep learning* têm demonstrado grande potencial na área do processamento de imagem, nomeadamente para a segmentação e classificação em imagens de ultrassom da mama. Contudo, ainda há necessidade de investigação e melhoria para ser possível aplicar com confiança estas abordagens na prática clínica. Adicionalmente, o ultrassom pode ser utilizado para orientar a agulha durante a biópsia mamária, um procedimento que exige elevado rigor.

O projecto OncoNavigator, no âmbito do qual esta dissertação foi realizada, visa combinar imagem médica em tempo real com um robô médico colaborativo para melhorar a precisão do rastreio do cancro da mama e na biópsia mamária guiada por ultrassom.

Considerando a necessidade de métodos automáticos para a segmentação e classificação para melhorar a prática clínica e de modo a aplicar inteligência artificial no robô médico colaborativo, nesta dissertação, foi desenvolvida uma nova *multi-task learning network* para a segmentação e classificação simultânea de tumores em imagens de ultrassom de mama. O método proposto foi avaliado em 810 imagens de ultrassom de dois conjuntos de dados, o BUSI e o UDIAT, tendo obtido um *Dice* de 80.72%, na segmentação, e *area under the curve* de 94.34%, na classificação. Em suma, o método demonstrou ser bem sucedido no delineamento e categorização de lesões e revelou potencial para ser incorporado num robô médico colaborativo para intervenções de cancro da mama e para auxiliar no seu diagnóstico.

Palavras-chave: Cancro da Mama, Ultrassom, *Deep Learning*, *Multi-Task Learning*, Segmentação de Tumores, Classificação de Tumores, *Computer-Aided Diagnosis*.

Abstract

Breast Tumor Segmentation and Classification using Deep Learning Methods

Breast cancer is the most common type of cancer and the leading cause of cancer death in women. Early detection is crucial for successful treatment and for reducing the mortality rate. Although several imaging modalities are used to detect breast lesions, ultrasound has become one of the most frequent as it is safe, portable, cost-effective and allows real-time examination. Thus, automatic tumor segmentation and classification in breast ultrasound images may aid their diagnosis, providing a second opinion to the expertise. However, accurate segmentation and classification of lesions using this imaging modality is challenging due to the poor image quality and high lesion variability. Recently deep learning algorithms have shown great potential in the image processing field, namely for breast ultrasound segmentation and classification. Nevertheless, there is still a need for research and improvement to confidently apply these approaches in clinical practice. Additionally, ultrasound can be used for needle guidance during breast biopsy, a procedure that requires high accuracy.

The OncoNavigator project, under which this dissertation was conducted, aims to combine real-time medical imaging with a collaborative medical robot to improve the precision of breast cancer screening and ultrasound-guided breast biopsy.

Considering the need for automatic methods for segmentation and classification to improve clinical practice and to apply artificial intelligence to the collaborative medical robot, in this dissertation, a novel multi-task learning network for simultaneous tumor segmentation and classification in breast ultrasound images was developed. The proposed method was evaluated on 810 ultrasound images from two datasets, BUSI and UDIAT, and obtained a Dice of 80.72% in segmentation and area under the curve of 94.34% in classification. In summary, the method proved successful in delineating and categorizing lesions and revealed potential to be incorporated into a collaborative medical robot for breast cancer interventions and to assist in its diagnosis.

Keywords: Breast Cancer, Ultrasound, Deep Learning, Multi-Task Learning, Tumor Segmentation, Tumor Classification, Computer-Aided Diagnosis.

Contents

List of Figures

List of Tables

Acronyms

2Ai   Applied Artificial Intelligence Laboratory 13

2D   Two-Dimensional 6, 7, 31, 32, 43, 46, 57, 61, 64, 66, 72, 74, 75

3D   Three-Dimensional 6, 61, 75

AI   Artificial Intelligence 10, 13, 15, 16, 72, 73

ANN   Artificial Neural Network 10

BI-RADS   Breast Imaging Reporting and Data System 7, 9, 21, 45

CAD   Computer-Aided Diagnosis 10, 12, 17, 45, 60, 70

CNN   Convolutional Neural Network 11, 12, 17, 18, 19, 21, 22, 23, 31, 32, 37, 38, 42, 45, 46, 47, 48, 61, 63, 73

DL   Deep Learning 10, 12, 13, 15, 16, 17, 19, 20, 21, 23, 29, 31, 32, 39, 41, 43, 45, 46, 47, 54, 57, 60, 72, 74, 75

DM   Digital Mammography 5, 6, 7, 12

EMBC   Engineering in Medicine and Biology Conference 13, 31, 45, 72

FC   Fully Connected 11, 19, 61, 64

FCN   Fully Convolutional Network 18, 19, 60

GPU   Graphics Processing Unit 39, 53

ILSVRC   ImageNet Large Scale Visual Recognition Challenge 17, 46

ML   Machine Learning 10, 16, 17, 21, 39

MRI   Magnetic Resonance Imaging 5, 6, 7, 12

MTL   Multi-Task Learning 15, 18, 19, 23, 60, 61, 64, 65, 66, 67, 69, 70

ROI   Region of Interest 46, 52

SASYR   Symposium of Applied Science for Young Researchers 14, 31

US   Ultrasound 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 19, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 35, 37, 38, 41, 42, 43, 45, 46, 47, 51, 52, 53, 54, 56, 57, 60, 61, 64, 66, 69, 70, 71, 72, 73, 74, 75

<div align="right">

1

</div>

Introduction

In this chapter, the clinical contextualization and motivation of the driving problem of this thesis are presented, along with the definition of the objectives and contributions of the project. Finally, an overview of the present document is given, summarizing the main contents of each chapter.

1.1   Breast Anatomy

The breast is a superficial structure located on the chest wall between the second and sixth or seventh ribs, lateral to the sternum, and medial to the anterior axillary line [1].

The breast is composed by the glandular tissue and by the dense connective tissue contained within an abundance of adipose tissue (fat), which is surrounded by skin [2]. The different tissues and anatomical structures that constitute the breast are depicted in Figure 1. It varies considerably in shape, size, and density, and its volume is determined by the amount of adipose and glandular tissue. The stroma that surrounds the glandular tissue is composed of fat and connective tissue. The glandular tissue consists of 15 to 20 radially situated lobes, separated from each other by connective tissue. Each lobe is subdivided into numerous lobules, the mammary glands, which are connected to the nipple through lactiferous ducts. The 15 to 20 main ducts branch and terminate in the terminal duct lobular unit (or lobule) responsible for milk secretion during lactation [1]. At the apex of the breast is a pigmented area, the areola, surrounding a central elevation, the nipple [2].

The upper lateral region of the breast extends into the axilla, forming the axillary tail, which contains the axillary lymph nodes, where occurs the lymphatic drainage from the breast [3]. The breast can be divided in six sub-sites: the upper-outer quadrant; upper-inner quadrant; lower-outer quadrant; lower-inner quadrant; central portion; and nipple [4], which are illustrated in Figure 2. Breast tumors can arise in any of these regions, but are most often located in the upper outer quadrant and less common in the lower inner quadrant [4].
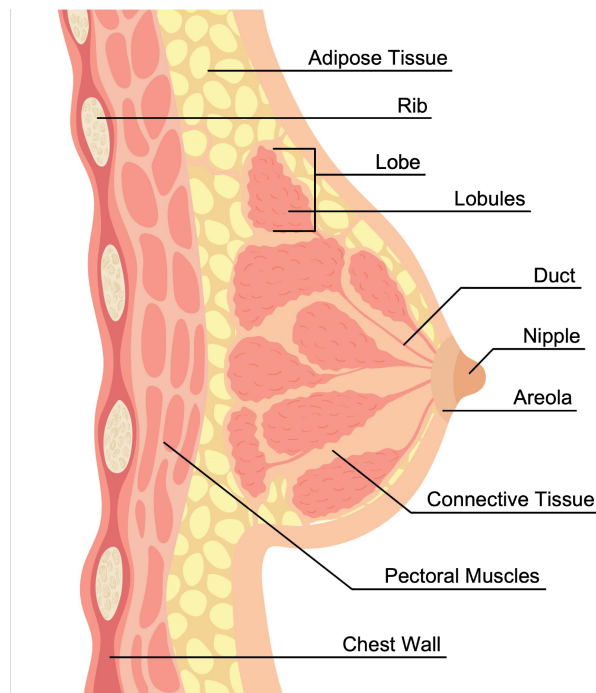
Figure 1: Anatomy of the breast (side view). Adapted from [5].

1.2   Breast Cancer

Breast cancer arises when abnormal breast cells grow and multiply in an uncontrolled manner, originating a mass of tissue termed tumor or neoplasm [6]. Breast cancer can emerge on different areas of the breast, namely the ducts, the lobules, and the tissue in between. However, most begin in the lobules or the ducts of the breast. In some cases, the tumor infiltrates the skin or components of the chest wall, such as the pectoral muscles. Moreover, tumor cells can convert the microenvironment into a favorable state to promote their growth and expansion [7].

Tumors are classified as benign (usually not life-threatening) and malignant (has the potential to be dangerous) [8]. Benign lesions are not cancerous, have a regular arrangement of the cells, and remain in their original location. In contrast, malignant masses are cancerous, rich in cellularity and cell clusters, and can spread to other tissues or organs [7, 9].

Besides the distinction between malignant and benign, breast cancer can be considered either invasive or non-invasive. When the cluster of cells remains within the basement membrane of the elements of the terminal duct lobular unit or the draining duct, it is categorized as *in situ* or non-invasive. In invasive breast cancer the dissemination of cancer cells occurs outside the basement membrane of the ducts and lobules into the surrounding adjacent normal tissue and can potentially disseminate to farther site (metastasis) [10].
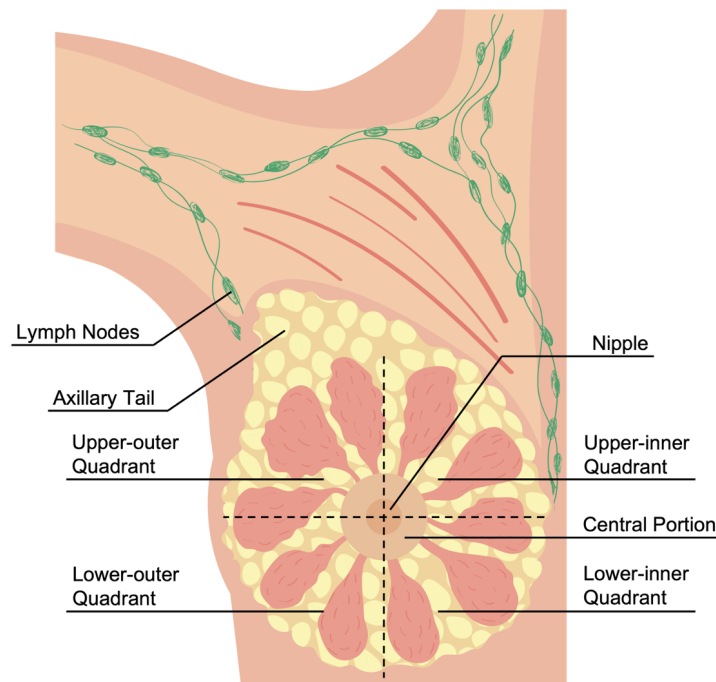
Figure 2: Schematic representation of the breast (front view of the right breast). Adapted from [5].
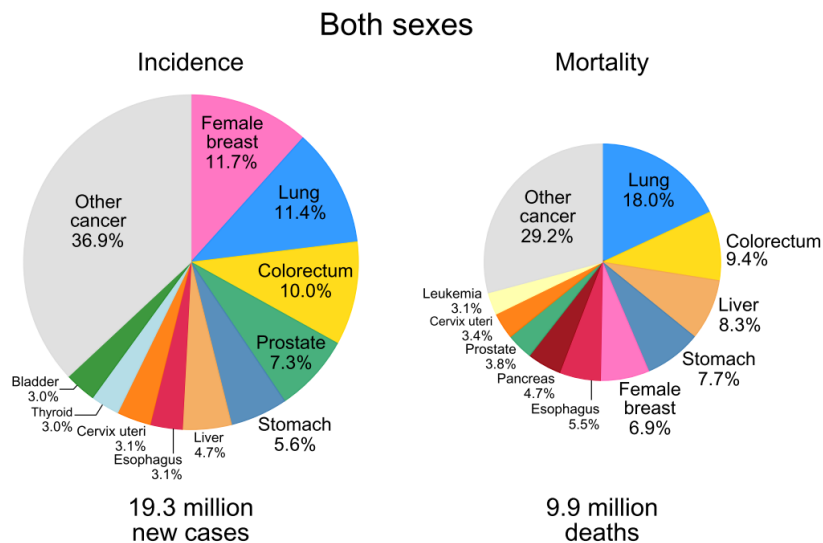
Figure 3: Distribution of cases and deaths for the top 10 most common cancers in 2020 for both sexes [11].

1.2.1 Epidemiology

Breast cancer is the most common cancer in the world, affecting mainly women and having the highest cancer mortality among them. Specifically, breast cancer is the most frequently diagnosed cancer with 11.7% of total cases, and the fifth leading cause of cancer death with 6.9% of the total cancer deaths (Figure 3). Moreover, in 2020, there were nearly 2.3 million (24.5%) newly diagnosed cases and 685,000 (15.5%) deaths from breast cancer [11] (Figure 4).
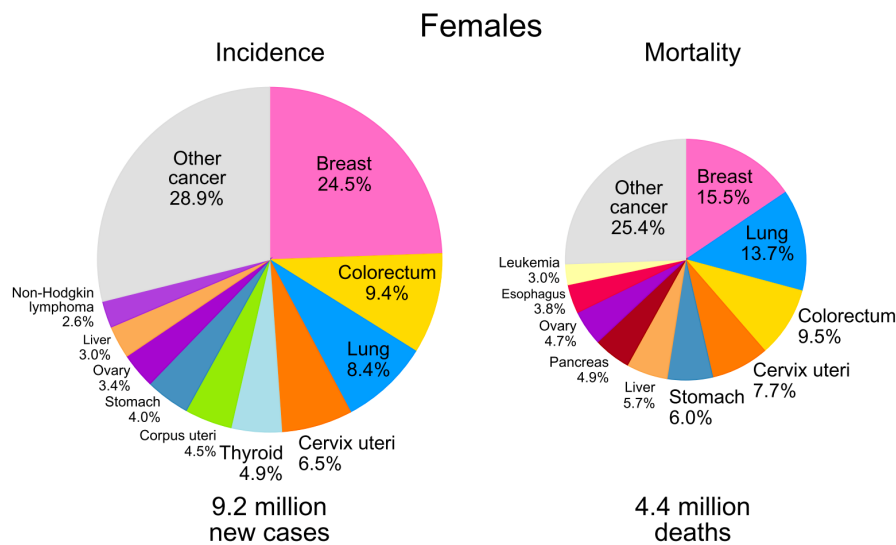
Figure 4: Distribution of cases and deaths for the top 10 most common cancers in 2020 for women [11].

The probability of a female being diagnosed with breast cancer has significantly increased from 1 in 11 women in 1975 to 1 in 8 [12]. Although breast cancer primarily emerges in women, an estimated 1% of cases are detected in men [13].

Despite the advances in breast cancer research, the cause of most cases is unknown. However, epidemiological studies have identified a wide range of risk factors for this disease, including race, ethnicity, increasing age, family history of cancer, genetic traits, and genetic mutations (namely the BRCA1 and BRCA2 genes), as well as increased alcohol consumption, physical inactivity, exogenous hormones, female reproductive factors and previous exposure to therapeutic chest wall irradiation [14, 15].

1.2.2   Types of Lesions

Breast lesions can be classified as benign or malignant (Table 1). Benign masses remain confined to their original location and do not invade nearby tissues or metastasize. They are generally surrounded by a fibrous capsule and tend to grow along with the structures of the breast. The most common types of benign lesions in the female breast are fibroadenomas, epithelial tumors containing fibrous tissue, and cysts, closed cavities in the body retaining a liquid or semi-solid material. With an incidence of 25%, fibroadenoma is the most frequent benign tumor, but treatment is usually not required [16]. Other benign breast anomalies include: fibrocystic breast disease (swelling of the breast due to fluid-filled cysts); duct ectasia (thickening of the duct walls); duct papillomas (small growth inside the duct); epithelial hyperplasia (overgrowth of the cells that line the lobules or ducts); intramammary lymph nodes; benign lymphomas (tumor that develops from lymphocytes) and mastitis (inflammation of breast tissue, with possible infection) [16, 17].

Malignant tumors tend to grow in irregular directions and can spread to surrounding tissue and metastasize through the body via the circulatory or lymphatic system. Breast carcinomas are categorized as

Table 1: Types of breast tumors.

Lesion Type	Diagnosis
Benign	— Fibroadenoma — Cyst — Fibrocystic breast disease — Duct ectasia — Duct papilloma — Epithelial hyperplasia — Intramammary lymph node — Benign lymphoma — Mastitis
Malignant	— Ductal carcinoma *in situ* — Invasive ductal carcinoma — Lobular carcinoma *in situ* — Invasive lobular carcinoma

ductal carcinomas if formed in the cells lining the ducts or as lobular carcinomas if formed in the lobules. Most breast cancer cases correspond to lobular carcinomas. Depending on whether the cancer cells invade nearby tissues, tumors may be classified as invasive (infiltrating) or non-invasive (*in situ*). Invasive carcinomas are the most prevalent form of breast cancer, in which cells spread outside the basement membrane of the ducts and lobules into the surrounding tissue. Infiltrating breast cancer include invasive ductal carcinomas and invasive lobular carcinomas. *In situ* breast cancer, although less aggressive, can eventually progress to invasive cancer and can be considered a potential cancer precursor. The two subtypes of non-invasive breast cancer are intraductal carcinoma *in situ* and lobular carcinoma *in situ* [7]. Hematological malignancies also affect the breast, lymphoma being the most common [18]. Illustrative examples of each type of lesion acquired using different imaging modalities are presented in the next section.

Calcifications are breast abnormalities that consist of calcium deposits in the breast, usually resulting from natural aging. Although these are often benign, they can be an early sign of breast cancer [19].

1.3 Breast Cancer Diagnosis

The early diagnosis and treatment of breast cancer has a decisive impact on the prognosis. Breast cancer diagnosis usually involves three sequential steps: physical examination, medical imaging examination, and biopsy [20]. Medical image examination is the most efficient way for breast cancer diagnosis. Different medical imaging modalities may be used during the diagnosis process, namely Ultrasound (US), Digital Mammography (DM), and Magnetic Resonance Imaging (MRI) [21].

(a) Benign mass   (b) Malignant mass
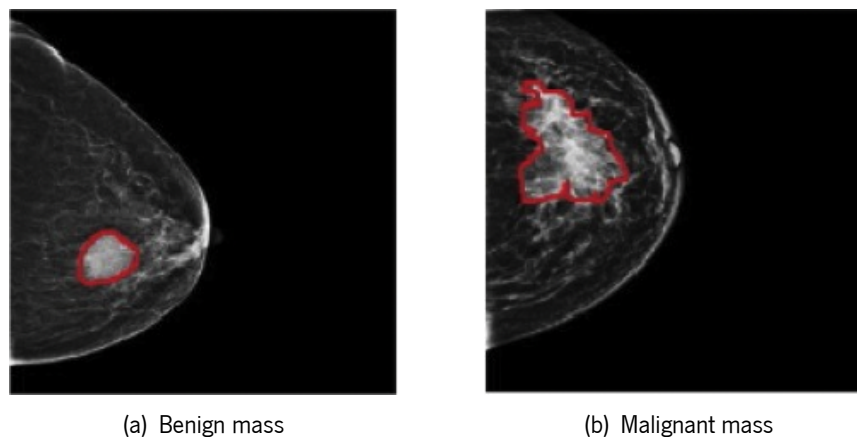
Figure 5: Examples of breast masses on mammograms. Adapted from [25].

1.3.1 Medical Imaging

For breast cancer diagnosis, medical imaging modalities, such as DM, MRI, and US, are used. DM is the most common technique for breast cancer detection, using low-dose X-rays to obtain high quality images and examine the breast [22]. DM is the standard breast cancer screening method, offering high Two-Dimensional (2D) resolution and being highly sensitive to the detection of calcifications. However, it has disadvantages, such as having low sensitivity for patients with dense breasts and using ionizing radiations [23]. As seen in Figure 5, masses appear brighter on the mammogram, as well as microcalcifications [24].

MRI is a widely used screening tool for breast cancer, providing a high resolution Three-Dimensional (3D) image (composed of cross-sectional images) of the breast using a strong magnetic field along with pulsing radio waves [23]. A contrast-enhanced breast MRI scan involves injecting a contrast dye into the patient to enhance abnormal areas [26]. MRI has a higher sensitivity than other imaging modalities, not being as affected by breast density as DM [27], and provides information about the vascularity of the breast tissue [8, 28]. However, it is expensive, time-consuming, and has a higher rate of false positives. Consequently, breast MRI is mainly recommended for screening women at high risk of breast cancer and evaluating therapy effects [23]. As illustrated in Figure 6, breast tumors appear brighter on MRI scans due to the accumulation of the contrast agent in the breast tumor stroma [29].

Another widespread tool for breast cancer detection is US, being relatively cheap, portable, real-time, non-radiation, and effective in detecting mammographically occult tumors in women with dense tissue [30, 31]. The US transducer directs high-frequency sound waves into the breast tissues and detects the reflected sound waves, which are used to display 3D or, more commonly, 2D images. As the sensor is moved over the breast, continuous real-time images can be captured [24]. However, US is highly operator-dependent, and its interpretation is challenging due to the inherent speckle noise, low signal-to-noise ratio, and shadows [32]. Currently, US has been recommended as an adjunctive modality to mammography, and is commonly used for guided biopsy examinations [8].
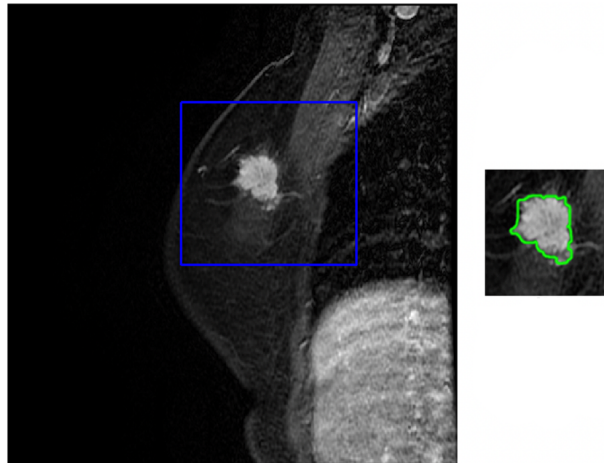
Figure 6: Example of a contrast-enhanced breast lesion on MRI. Adapted from [29].

Breast lesions in 2D US images can be described regarding their their textural characteristics (for example, echo pattern and posterior acoustic features), and morphological characteristics, such as shape and margin. The shape of the lesion can either be regular or irregular. Irregular shape is often related to malignant lesions. The margins of the tumors can be circumscribed (well-defined) or non-circumscribed (indistinct, angular, micro-lobulated or spiculated). Tumors with circumscribed margins are frequently benign, whereas tumors with non-circumscribed margins are indicative of malignancy. The boundary of the lesion can have a clear delineation between the lesion and the surrounding tissue, or have a echogenic halo with no strong demarcations. The last is frequently related to malignancy. The echo pattern of a tumor can be categorized as anechoic (no internal echoes), isoechoic (same echogenicity as fat), hyperechoic (increased echogenicity relative to fat; brighter appearance), or hypoechoic (low internal echoes relative to fat; darker appearance). Malignant tumors often have a hypoechoic or isoechoic echo pattern. Moreover, tumors with posterior acoustic shadowing are indicative of cancer [33]. A brief description of the breast US images with respect to their sonographic appearance details are given in Figure 7.

1.3.2 BI-RADS

The Breast Imaging Reporting and Data System (BI-RADS) lexicon, developed by The American College of Radiology, provides a standard for experts to interpret the medical images of breast cancer (DM, US, and MRI). This provides standard descriptions of breast imaging features used to classify lesions and ultimately present an overall risk for malignancy [34].

In this system, the image results are classified into BI-RADS categories numbered 0 through 6 [35]. A brief explanation is presented below:

- **Category 0** - *Incomplete.* Additional imaging evaluation and comparison to previous exams are needed. This means a possible abnormality may not be clearly seen, and more tests are required.
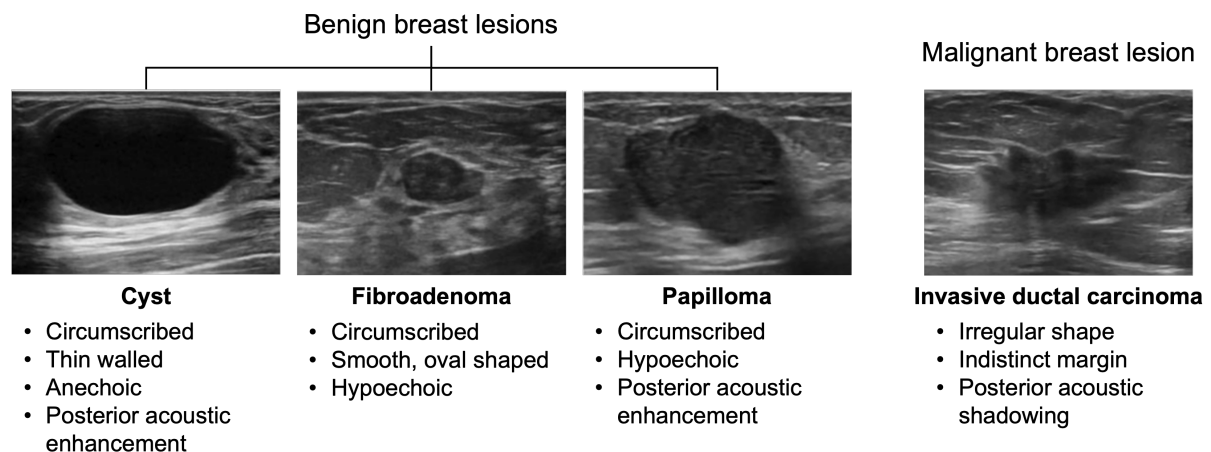
Figure 7: Examples of different breast lesions on US and respective descriptions regarding sonographic appearance. Adapted from [33].

The exam should be compared to other images, if available, to discover if the area has changed over time.

- **Category 1** - *Negative.* There is no significant abnormality to report. The breasts are symmetrical with breast lesions.

- **Category 2** - *Benign finding.* There is no sign of cancer (0% probability of malignancy), but a benign structure was found and reported to avoid misinterpretations.

- **Category 3** - *Probably benign finding.* A lesion was found but it is not proven benign and there is a less than 2% chance of malignancy. Therefore, follow-up in a short time frame is suggested to analyze possible changes over time. The findings in this category have a high probability (greater than 98%) of being benign (not cancer).

- **Category 4** - *Suspicious abnormality.* The findings are suspicious and could be cancer. Therefore, biopsy should be considered. There are sub-categories depending on the suspicion of malignancy:

  — **Category 4A** - Low suspicion for malignancy (2% to 8% probability of malignancy).

  — **Category 4B** - Moderate suspicion for malignancy (9% to 49% probability of malignancy).

  — **Category 4C** - High suspicion for malignancy (50% to 95% probability of malignancy).

- **Category 5** - *Highly suggestive of malignancy.* The findings appear to be cancer and have a chance of 95% or more of being cancer. Thus, biopsy is strongly recommended.

- **Category 6** - *Known biopsy-proven malignancy.* This category is only used for findings that have already been proven to be cancer by a previous biopsy [35].
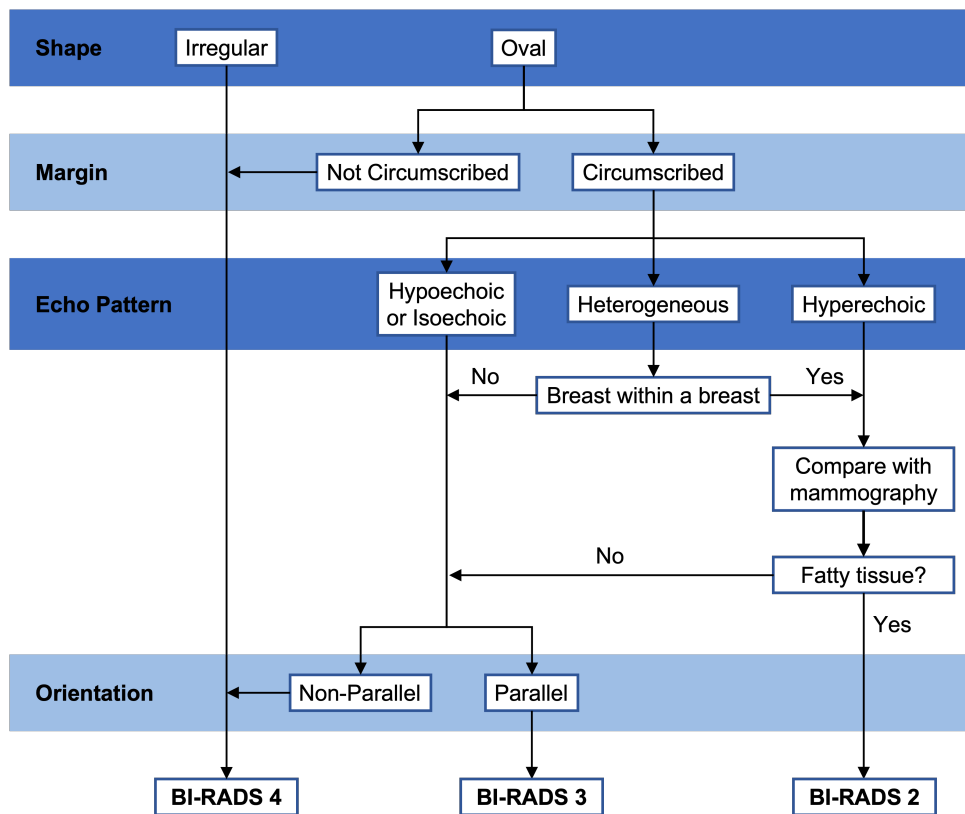
Figure 8: Flowchart for breast tumor analysis using US. Adapted from [36].

Figure 8 shows a flowchart for breast tumor classification using US. To classify a breast lesion into BI-RADS categories, the first step is to analyze the shape of the lesion. If it is irregular in shape, it is categorized as suspicious (Category 4). If the lesion has a regular shape, it is necessary to assess the margin [36].

When the margin is non-circumscribed, namely angular, indisctinct, spiculated or microbulated, even if in a small portion of the lesion, then it is categorized as BI-RADS 4. Otherwise, the margin is circumscribed and the echogenecity of the tumor must be evaluated [36].

When a tumor is hypoechoic or isoechoic, the orientation dictates if it is suspicious or probably benign (Category 3). It is classified as BI-RADS 4, if it is non-parallel to the skin, and categorized as BI-RADS 3, when parallel to the skin. If the lesion is heterogeneous and the background homogeneous, it is suspicious if it is not parallel to the skin, and classified as probably benign if it is parallel. When a lesion is heterogeneous (as well as its surroundings) or hyperechoic more tests are required, such as mammography. Then, the presence of fat is investigated. If the lesion contains fat, it is classified as benign (Category 2). On the other hand, if there is absence of fat, the tumor is either classified as BI-RADS 3 or 4, depending on the orientation [36].

9
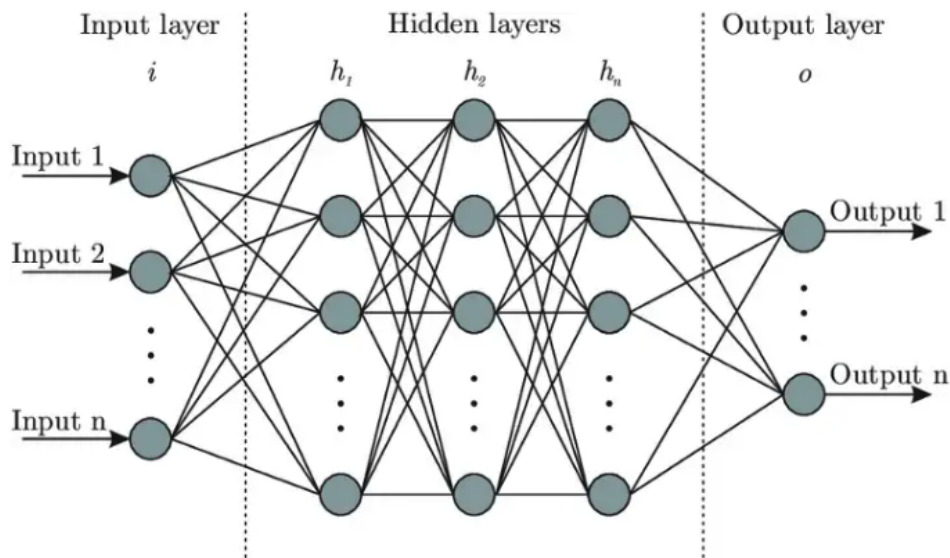
Figure 9: Neural network structure [41].

1.3.3 Computer-Aided Diagnosis

Breast cancer diagnosis through medical images requires analysis by an expert radiologist, which is time-consuming and prone to mistakes [19]. However, particularly for US, image interpretation can be difficult due to the inherent speckle noise, low signal-to-noise ratio, and shadows [37]. Computer-Aided Diagnosis (CAD) systems may help overcome these challenges, assisting radiologists in the detection, segmentation, and classification of masses in breast images. The main goal is to improve the diagnosis accuracy by providing specialists a second opinion [19]. The workflow of a CAD system can be divided into several steps, such as data preprocessing, image segmentation, and image classification [38]. In CAD systems, an algorithm performs the processing and analysis of the medical image and determines the lesion characterization, potentially resulting in a more objective and accurate diagnosis [21]. These systems can be used not only during the diagnostic process, but also during breast cancer interventions, such as breast biopsy.

Different Artificial Intelligence (AI) algorithms are gaining attention due to their excellent performance in image recognition tasks [19]. Machine Learning (ML) algorithms are being suggested as an alternative to human vision and experience for the segmentation and classification of breast lesions. Moreover, Deep Learning (DL), a sub-field of ML, has made significant progress in medical imaging, revealing the great potential for image analysis tasks [23, 39]. Neural networks, also termed Artificial Neural Networks (ANNs), are the foundation for DL algorithms, and their structure and nomenclature are inspired by the biological neural networks that constitute human brains [40]. ANNs comprise several node layers, containing an input layer, one or more hidden layers, and an output layer (Figure 9).

Each node (artificial neuron) connects to another and has an associated weight and threshold. If the output of a node is above the specified threshold value, that node is activated, sending data to the
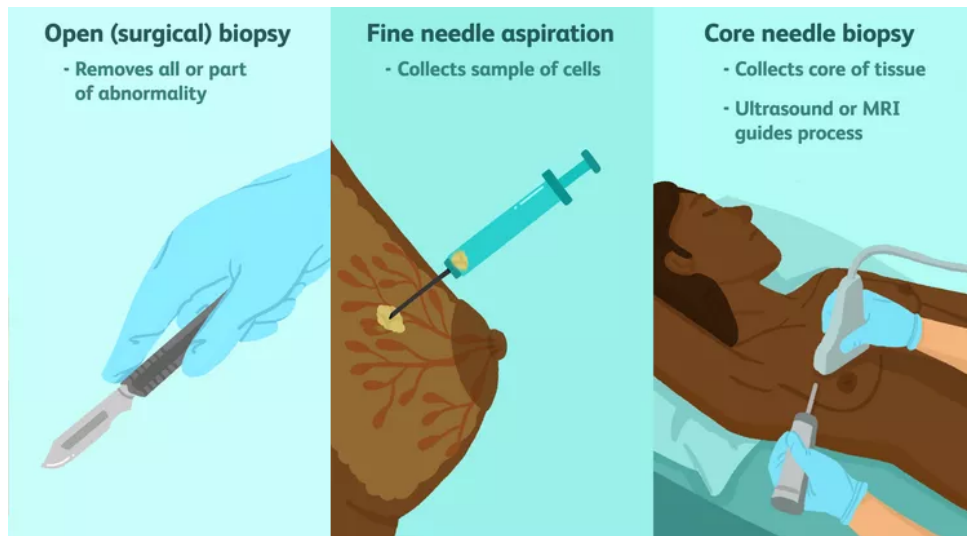
Figure 10: Types of breast biopsies. Adapted from [45].

next layer of the network, otherwise, no data passes to the next layer of the network [40]. The data is passed through several layers, and each layer extracts features progressively and passes them to the next layer. Initial layers extract low-level features, and succeeding layers combine features to form a complete representation. In supervised learning, labeled data is used to train the neural network, whereas, in unsupervised learning, the network uses unlabeled data and learns the recurring patterns [42].

Studies have shown that Convolutional Neural Networks (CNNs) outperform medical experts in certain tasks [43]. CNNs are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs and constitute three main types of layers: convolutional layer, pooling layer, and Fully Connected (FC) layer.

1.3.4 Biopsy

Even considering recent advances in breast imaging, a variety of benign abnormalities can resemble cancer, and vice versa, in all imaging modalities. Therefore, biopsy represents the most accurate method to confirm or exclude malignancy [44]. A biopsy is an invasive procedure where a tissue sample of a breast lesion is extracted to diagnose [28]. Figure 10 summarizes and illustrates the most frequent types of breast biopsies.

Biopsies can be surgical or percutaneous. A surgical biopsy involves the removal of all (excisional biopsy) or part (incisional biopsy) of a suspicious area to check for cancer cells [44]. Percutaneous image-guided needle biopsy has replaced surgical biopsy allowing a minimally invasive, safe, accurate, and cost-effective diagnosis of breast lesions necessary for treatment planning [44]. The most common technical options for percutaneous biopsies are fine-needle sampling, which provides material for studying cells (cytological examination), and core needle biopsy, which provides material for studying tissues (histopathological examination) [44]. In fine-needle sampling, a fine needle is inserted close to the US

probe, and when it reaches the target, manual multidirectional sampling is performed through aspiration using a syringe or a vacuum aspiration system [44]. A core needle biopsy involves the insertion of a larger needle, often through a small skin incision, and once it is confirmed to be on target, tissue samples (cores) are obtained [44].

All these techniques can be conducted using US, DM, or MRI to guide the biopsy needle to the target area of the tissue to be sampled and examined [46]. If the lesion is visible on US, this technique is preferred due to the simple approach, the short duration of the procedure, and the lower cost. In case the lesion is not identifiable on US, other imaging modalities are used for guidance, typically DM for suspicious calcifications and architectural distortions and MRI for lesions that are not visible on US and DM [44].

Although biopsy has been considered the gold standard for breast cancer diagnosis, the rate of positive findings at biopsy is significantly low [20]. Therefore, the implementation of automatic techniques using non-invasive imaging modalities can improve the diagnostic accuracy and reduce the number of unnecessary biopsies.

1.4   Motivation

As abovementioned, breast cancer is the most prevalent cancer among women and one of the leading causes of death worldwide [11]. It currently affects one in eight women globally, and its incidence and mortality have increased year by year [28]. Early detection of breast cancer allows an improved prognosis and helps reduce mortality rates as more efficient treatments can be provided.

US has become one of the most common imaging modalities for breast cancer diagnosis. In traditional clinical practice, breast US images may require manual segmentation and classification of the lesions by an expert radiologist, which is time-consuming, prone to errors, and observer-dependent [47]. Furthermore, these tasks are hampered not only due to the US poor image quality but also due to the high variability of lesions in shape, size, texture, and location [48].

With the emergence of DL, medical image interpretation has substantially been improved through CAD systems using CNNs. Although many automatic methods for breast lesion segmentation and classification in US images have been proposed, there is a continuing need for more refined architectures and more extensive datasets [49].

In addition to US being used as a diagnostic tool, it can be used for needle guidance during breast biopsy, a procedure that requires high accuracy. The OncoNavigator project proposes a novel navigation framework that combines the information from real-time medical imaging with a collaborative medical robot. The main goal of this project is to improve the precision of current breast cancer interventions, namely the US-guided breast biopsy, while guaranteeing a safe, faster, and easier treatment. The proposed framework will assist medical professionals during breast biopsies using an US probe attached to a collaborative robot arm along with a sensorized guide to lead the biopsy needle to the target lesion. In addition, a user interface will allow visualizing the real-time image acquired through the US probe, and an

AI module will segment and classify the lesions in the image. A trajectory from the biopsy needle is also displayed to help the user in the procedure. Thus, when the lesion segmentation option is active, it is possible to check whether the needle trajectory is correct. Therefore, the proposed framework will consist of the following modules: collaborative robot arm, portable US scanner system, sensorized needle tracker, and the automatic breast lesion segmentation and classification module.

This project is a part of the Applied Artificial Intelligence Laboratory (2Ai), a research-based organization within the Polytechnic Institute of Cávado and Ave. This dissertation was carried out within the OncoNavigator project and focused on the development of a breast tumor segmentation and classification method for US images to be incorporated in the OncoNavigator framework, as well as to assist in breast cancer diagnosis. Other members are responsible for building a sensorized needle guide, developing the user interface and software to integrate all the framework modules, and designing the mechanical components.

1.5 Aims and Contributions

The focus of this work is to develop a breast tumor segmentation and classification method for US images using DL. The primary goal of this algorithm is to accurately delineate and classify a possible lesion in a breast US image for an objective, correct, and efficient diagnosis. The proposed segmentation and classification method is twofold, as it is intended to be used to assist radiologists in breast cancer diagnosis, as well as integrated into a collaborative medical robot for US-guided breast biopsies to improve the efficiency and accuracy of these interventions.

To this extent, the main contributions of this work are as follows:

- Collection of public breast US datasets for the construction of an extensive dataset (**Chapter 3**);

- Review and comparative evaluation of state-of-the-art deep neural networks for breast lesion segmentation (**Chapter 4**);

- Review and comparative evaluation of state-of-the-art deep neural networks for breast lesion classification (**Chapter 5**);

- Development of a new DL architecture for simultaneous lesion segmentation and classification in breast US images (**Chapter 6**).

An overview of the steps of the project, that resulted in the abovementioned contributions, is presented in Figure 11.

One of the contributions, corresponding to Chapter 4, was accepted and published in the proceedings of the Engineering in Medicine and Biology Conference (EMBC) 2022: M. R. Ferreira *et al.*, "Comparative Analysis of Current Deep Learning Networks for Breast Lesion Segmentation in Ultrasound Images,"*2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*,
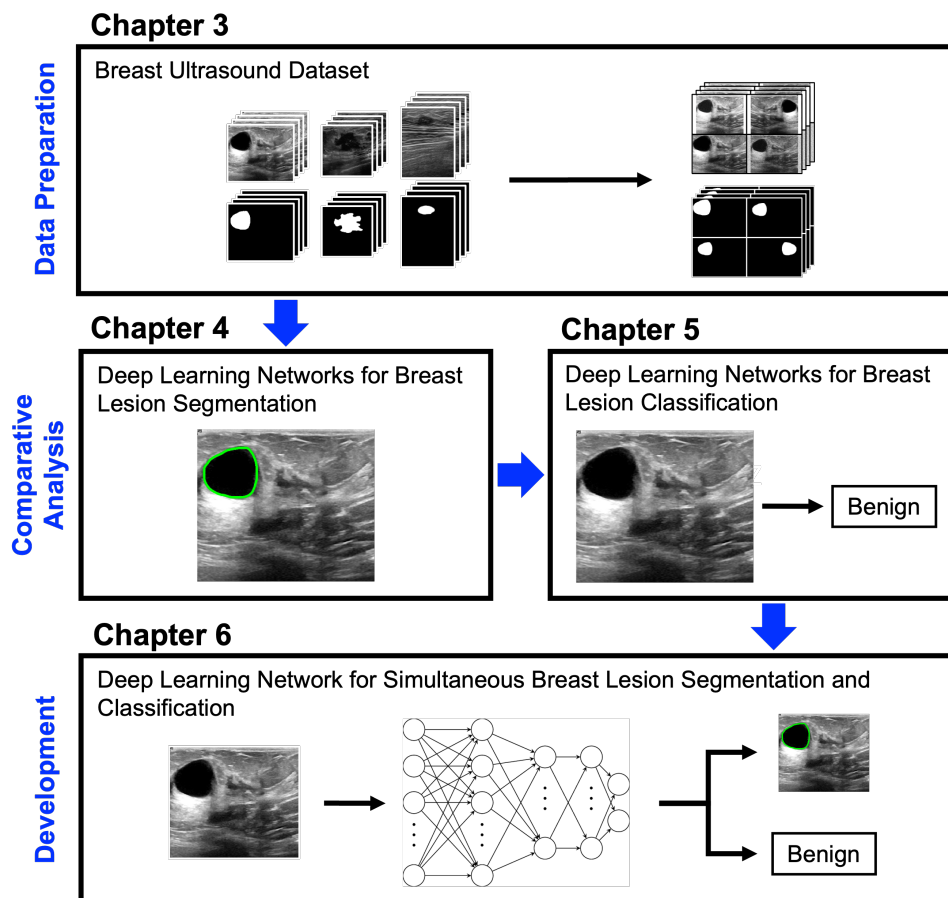
Figure 11: Overview of the project steps.

2022, pp. 3878-3881, doi: 10.1109/EMBC48229.2022.9871091. Moreover, an abstract on this work was accepted and published in the book of abstracts of the Symposium of Applied Science for Young Researchers (SASYR) 2022: M. Ferreira, H. Torres, B. Oliveira, P. Morais, P. Novais, and J. Vilaça, "Deep Learning Networks for Lesion Segmentation in Breast Ultrasound: A Comparative Analysis," in *2nd Symposium of Applied Science for Young Researchers*, Bragança, 2022, pp. 6–8. Finally, the contributions presented in Chapters 2, Chapters 5 and 6 were documented as three research papers, ready for submission.

1.6 Research Questions

According to the contributions sough in this thesis, the following research questions were identified and answered:

- **RQ1** - What are the segmentation and classification models implemented in scientific literature for breast lesions using US images?

- **RQ2** - Which segmentation and classification models tested on a large dataset with the same evaluation metrics perform better?

- **RQ3** - What is the influence of using different input data variations, that incorporate segmentation information, for breast lesion classification models?

- **RQ4** - Is it possible to improve the performance of segmentation and classification tasks by using Multi-Task Learning (MTL) strategies?

- **RQ5** - Are the performances of the DL models adequate and comparable to human evaluation?

1.7 Dissertation Overview

This dissertation is organized into seven chapters. In the current chapter, an introduction to the study was established. The anatomy of the breast was presented, along with a contextualization of breast cancer and its diagnostic process. Furthermore, the motivation, the aims, and the contributions were defined.

Chapter 2 addresses fundamental ideas of AI and anomaly detection, as well as medical image segmentation and classification. Moreover, a review of state-of-the-art DL methods for breast lesion segmentation and classification in US images is presented.

In Chapter 3, the used breast US datasets were detailed, along with the procedure followed to construct and prepare the final dataset for the study.

Chapter 4 describes a comparative study of seven state-of-the-art DL networks for breast segmentation using the constructed dataset of breast US images. Similarly, Chapter 5 reports the evaluation and comparison of six current DL networks for breast classification using five image variations of the specified dataset. Both chapters include the methodology, obtained results, and a discussion.

In Chapter 6, the development of a DL network for simultaneous breast lesion segmentation and classification in US images is described. The network architecture is presented along with its performance on the constructed breast US dataset. The proposed method is validated through a comparison with other methods.

Finally, Chapter 7 presents the answers to the specified research questions and main conclusions regarding the developed work, outlining the main contributions and limitations. Moreover, some perspectives about future improvements and future work are given.

$$2$$

State-of-the-Art

This chapter provides an overview of AI methods proposed over the years for medical image analysis tasks, namely medical image segmentation and classification. Moreover, DL state-of-the-art approaches for breast tumor segmentation and classification in US images, introduced after 2016, are presented, along with their performances and used datasets. Methods to perform these tasks individually and simultaneously are reviewed, and their limitations are addressed. A review paper on the topics listed above has been written and prepared for submission.

2.1 Artificial Intelligence and Anomaly Detection

AI is commonly defined as the ability of a system to accurately interpret external data, learn from such data, and apply that learning to accomplish specific objectives and tasks through flexible adaptation. Over the last 50 years, the significant advances in computer power associated with big data emergence have pushed AI applications into new fields [21]. Currently, AI may be found in a wide variety of areas, such as robotics, image segmentation, image classification, speech recognition, natural language processing, data mining, and anomaly detection [50].

Anomaly detection is a broadly studied problem within diverse research domains, such as cybersecurity, manufacturing fraud detection, and medical imaging [51]. The main aim of anomaly detection is to identify data samples that do not fit the overall data distribution. Fundamentally, anomaly detection methods need to model the patterns in normal data to recognize atypical samples [51].

ML is a branch of AI that systematically applies algorithms to synthesize the underlying relationships among data and information [52]. Furthermore, DL, a sub-field of ML, is becoming increasingly popular among researchers as it offers many advantages over ML. DL increases the non-linearity of the model, allowing it to better differentiate between normal and inconsistent samples and model data discrepancies. Another advantage of DL is its capability to learn new features automatically. The emergence of big data and improved computational resources has empowered the hierarchical feature learning process that DL allows, overcoming the need for hand-crafted features used extensively in ML. Furthermore, DL can uncover long-term relationships within the data through the neural network architecture without explicitly stating

them [21, 50].

Anomaly detection is applied to various stages of medical data processing with DL, namely diagnosis through medical image segmentation and classification [50].

2.2 Medical Image Segmentation and Classification

Segmentation and classification are two fundamental steps in medical image analysis and CAD systems, playing a crucial role in the diagnosis of several diseases [21, 53]. Medical image segmentation is the process of extracting a relevant region from a medical image to help in the characterization of the tissue, which is widely employed for segregating the tumor regions from the background [54]. Furthermore, medical image classification aims to distinguish medical images according to a specified criterion, such as clinical pathologies or imaging modalities [55]. Often this task focuses on classifying a previously identified part of the medical image (obtained after a delineation task) into two or more classes, being used, for example, to categorize a tumor as benign or malignant [43]. However, both segmentation and classification usually require interpretation by an expert radiologist, which is time-consuming and prone to inter and intra-observer variability [23, 56]. Therefore, it is essential to develop advanced automatic medical image analysis methods to assist in diagnosis and improve its accuracy and reliability [43, 56].

Many researchers have proposed automated segmentation and classification algorithms for image analysis. Figure 12 shows a timeline of the proposal of relevant DL networks for segmentation and classification. Earlier approaches were built on traditional methods, such as edge detection filters and mathematical methods. ML approaches that require extraction of hand-crafted features were also proposed. In recent years, owing to hardware improvement and data availability, DL methods, especially CNNs, emerged and demonstrated their considerable capabilities in image processing tasks, reaching expert levels [57–59].

DL gained significant attention when, in 2012, Krizhevsky *et al.* proposed AlexNet, a CNN for image classification, which competed in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [60]. Two years later, GoogLeNet and VGGNet were proposed, improving the image classification accuracy on the ImageNet dataset [61, 62]. GoogLeNet was later refined as InceptionV3, presenting lower error and computational cost [63]. To increase the robustness of the input data, He *et al.* presented a spatial pyramid pooling model termed SPP-net [64]. The same authors later proposed the residual network, ResNet, for the problem of model degradation, which had a great influence on DL evolution [65]. In 2017, Huang *et al.* introduced the Dense Convolutional Network (DenseNet) with dense blocks that directly connect any two layers [66]. In 2018, Sandler *et al.* proposed the MobileNetV2, a CNN based on an inverted residual structure with bottlenecking features [67]. Both DenseNet and MobileNetV2 were evaluated on various object recognition benchmark tasks, including ImageNet, and obtained significant improvements over the state-of-the-art on most of them [66, 67]. In 2019, EfficientNet was proposed along with a new compound scaling method, achieving higher accuracy and efficiency than previous CNNs [68].

17
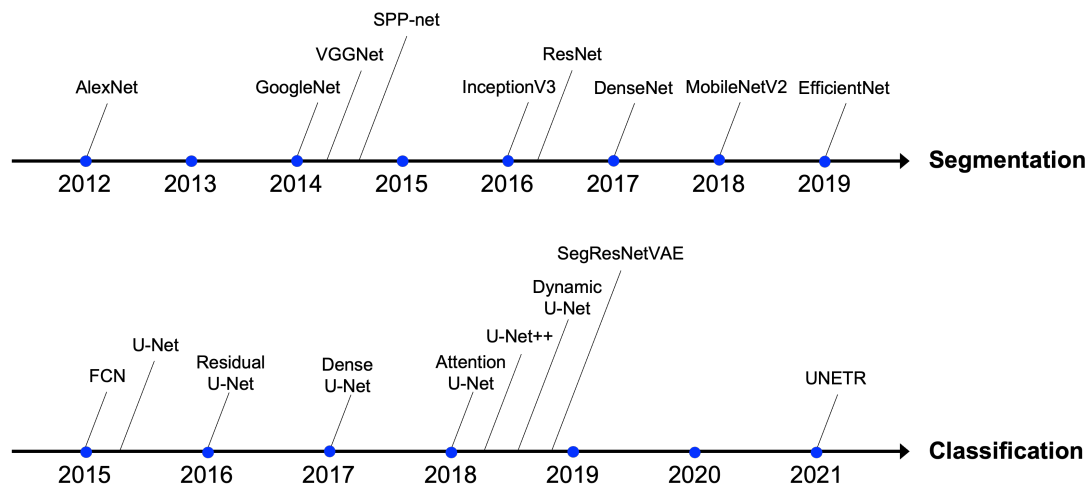
Figure 12: A timeline view of different state-of-the-art DL approaches for segmentation and classification.

Following CNN's accomplishments in classification, researchers employed CNN for image segmentation. In 2015, Long *et al.* adapted contemporary classification networks, AlexNet, the VGGNet, and GoogLeNet, creating the Fully Convolutional Network (FCN) for the semantic segmentation task [69]. In the same year, Ronneberger *et al.* proposed the U-Net, a pixel-to-pixel, end-to-end, FCN for medical image segmentation. Based on the idea of FCN deconvolution to restore image size and feature, U-Net presents the encoder-decoder architecture. The encoder is responsible for learning global contextual representations by gradually downsampling the extracted features, whereas the decoder is responsible for pixel or voxel-wise semantic prediction by upsampling the derived representations to the input resolution. In addition, skip connections combine the encoder's and decoder's outputs at multiple resolutions, allowing for the recovery of lost spatial information [70].

Recently, many U-Net-like architectures have been proposed, such as Residual U-Net [65], Attention U-Net [71], Dense U-Net [66], U-Net++ [72], and Dynamic U-Net [73]. In 2018, Myronenko *et al.* proposed SegResNetVAE, a residual U-Net with a variational autoencoder branch that provides a regularization effect [74]. Moreover, in 2021, Hatamizadeh *et al.* proposed U-Net Transformers (UNETR), which replaces the U-Net encoder with a Vision Transformer [75] generalization, to learn sequence representations of the input volume and effectively capture the global multi-scale information [76]. UNETR achieved a new state-of-the-art performance for many medical image segmentation tasks.

Since segmentation and classification tasks are highly related and share common image features (for example, a lesion boundary), solving these two tasks jointly in one unified model is a promising direction [77]. This technique implements MTL, a learning paradigm that aims at learning multiple related tasks in parallel and improves generalization abilities by sharing learned representations [77]. Therefore, what is learned for each task can help the model learn others better. MTL have been employed in many medical image analysis applications. Recently, a similar study that focuses on the segmentation and classification of tumors in mammograms has been presented in [78]. The proposed network combines

segmentation and classification using a FCN with a ResNet backbone to extract local features to jointly optimize both tasks [78]. Similarly, in 2021, Zhou *et al.* proposed a MTL framework for segmenting and classifying lesions in 3D automated breast US images [79]. In this network, a classification branch is added at the bottom of an encoder-decoder style segmentation network to enable the model to learn the two related tasks simultaneously. It has been proven that learning lesion classification and segmentation jointly can boost the performance of both tasks [79]. In 2022, Kang *et al.* proposed a thyroid nodule segmentation and classification network in US images that employs multi-stage MTL. It performs three tasks in two stages simultaneously: binary segmentation and classification in the first stage, and multi-class segmentation in the second stage. Additionally, intra and inter-task consistency is implemented to learn consistent predictions for all the tasks. For the segmentation tasks they used U-Net. After the encoder of the U-Net of the first stage, a classification branch is added, which consists of one adaptive average pooling layer, one flatten layer, one FC layer and one softmax layer. They concluded that using MTL and intra and inter-task consistent learning can improve the performance of all tasks for thyroid nodule segmentation and classification.

2.3 Deep Learning for Breast Lesion Segmentation in Ultrasound Images

To accurately characterize breast lesions, segmentation may be advantageous. The segmentation process of breast US images is characterized as the delineation of the lesion contours [21]. This process is crucial for accurate measurement of lesion size and shape. In early research, the main segmentation approaches included threshold-based, region-growth-based, and active contour-based methods [19]. These approaches often need extra manual intervention or a parameter selection process, which still leads to inter and intra-observer variations [19]. Recently, DL algorithms, particularly CNNs, have superseded the image processing field, showing promising results for breast US segmentation [80].

For this purpose, in 2019, Zhuang *et al.* proposed a Residual Dilated Attention Gate U-Net model (RDAU-Net) based on the conventional U-Net. RDAU-Net modified U-Net by implementing: residual units to enhance the edge information, Attention Gate modules to enhance the learning capabilities through suppression of background information, and dilated convolutions to increase the receptive field and acquire more characteristic information [81]. Negi *et al.* extended the work of Zhuang *et al.* [81] and proposed WGAN-RDA-UNET. A Wasserstein Generative Adversarial Network (WGAN) was employed, which comprises two modules, namely the generator and discriminator. RDAU-Net is used as the generator, and a CNN classifier is employed as the discriminator. The inclusion of a GAN-based approach improved the performance considerably [82]. In 2020, Shareef *et al.* presented a Small Tumor-Aware Network (STAN) to improve the segmentation of different-sized lesions. The STAN has two encoder branches that extract and integrate rich context information and high-resolution image features [37]. Although this model outperformed other

DL approaches, an Enhanced Small Tumor-Aware Network (ESTAN) was proposed to increase segmentation performance on small tumors [83]. ESTAN extends STAN and introduces row-column-wise kernels in the encoder to adapt to the breast anatomy and three skip connections in the decoder from the two encoders to combine features. In [84], a contextual information-aware conditional generative adversarial learning framework was proposed. An atrous convolution block is adopted to capture spatial and scale context to handle different tumor sizes and shapes. Moreover, channel attention and channel weighting mechanisms are employed to enhance relevant features and mitigate the effect of artifacts, as well as a combination of structural similarity index metric and L1-norm in the loss function to capture the local context information derived from the region surrounding the tumors [84]. In [85], a Second-Order Subregion Pooling Network (S2P-Net) is proposed, introducing an attention-weighted subregion pooling module to refine features by aggregating global features from the whole image and local information of subregions. In addition, a guided multi-dimension second-order pooling block is applied in each subregion to acquire powerful second-order covariance representations [85]. In 2021, several approaches were introduced, such as Global Guidance Network (GG-Net), which includes a global guidance block that aggregates non-local features under the guidance of multi-layer integrated features to learn powerful non-local contextual information [32]. Xu *et al.* proposed a Multi-Scale Self-Attention Network (MSSA-Net) that combines rich local features with global contextual information at many scales and applies self-attention to multi-scale feature maps [86]. Iqbal *et al.* presented a Multi-Scale Dual Attention-based Network (MDA-Net) for concurrent segmentation of breast lesions images. MDA-Net contains a multi-scale fusion block with various receptive fields to overcome fixed receptive field difficulties and extract semantic feature maps to achieve greater feature variability [87]. In [88], a Residual Cross-Spatial Attention-Guided Inception U-Net (RCA-IUnet) was introduced, based on the U-Net model with residual inception depth-wise separable convolution and max and spectral pooling layers. Furthermore, cross-spatial attention filters are added to suppress the irrelevant features and focus on the target structure [88]. Lastly, Wang *et al.* proposed a novel Residual Feedback Network (RF-Net) based on encoder-decoder architecture to optimize breast lesion segmentation by learning residual representations of hardly predicted pixels and feeding them into encoder blocks to increase the confidence [80].

Table 2 shows the referred methods along with the respective used datasets and results obtained for different evaluation metrics, namely, Dice Coefficient (DC), Jaccard Index (JI), accuracy (*Acc*), sensitivity (*Sen*), specificity (*Spe*), precision (*Pre*), F1-Score (*F1*), Area Under Curve (AUC), Precision-Recall Area Under Curve (PR-AUC), Receiver Operating Characteristic Area Under Curve (ROC-AUC), False Positive Ratio (FPR), Area Error Ratio (AER), Dice Loss (DCL), Average Hausdorff Error (AHE), Mean Absolute Error (MAE), Average Boundary Distance (ABD), Hausdorff Distance (HD) and Average Surface Distance (ASD).

Although the abovementioned networks reveal high breast lesion segmentation performances, there is still potential for improvement. Moreover, most networks were evaluated on different datasets, some being small or private, using different quantitative evaluation metrics, which makes it difficult to identify which network is better for this task. Therefore, there is a need to compare these approaches using the same dataset with a larger number of images and using a standardized evaluation method to determine which

segmentation network currently has the best performance and comprehend the architecture characteristics that contribute to this.

2.4 Deep Learning for Breast Lesion Classification in Ultrasound Images

The classification of breast lesions as benign or malignant is critical for determining the appropriate course of treatment. Radiologists mainly classify the lesions in breast US images based on BI-RADS. To make this assessment more objective and accurate, DL methods for breast lesion classification have been developed [21]. The main focus of most of the studies addressed in this section are related to the binary classification of benign against malignant lesions [30].

In 2017, Han *et al.* employed GoogLeNet [61] and modified this network for breast US image classification. The input layer was modified to deal with grayscale images, and since their purpose was to differentiate benign and malignant lesions, the output was reduced from 1000 to two class outputs [89]. In [90], three transferred models, namely InceptionV3 [63], ResNet50 [65], and Xception [91], a CNN model with three convolutional layers (CNN3), and a traditional ML-based model with hand-crafted features were compared. The transfer learning method outperformed the conventional ML model and the CNN3 model, with the transferred InceptionV3 achieving the best results [90]. In 2019, Byra *et al.* adapted VGG19 [62] for breast mass classification and introduced the concept of the matching layer, a layer used to convert grayscale US images to RGB. The authors demonstrated that the matching layer allows a more efficient classification compared to the duplication of grayscale US images across the RGB channels [92]. Fujioka *et al.* used GoogleNet InceptionV2, an improvement over GoogLeNet, to discriminate between benign and malignant breast masses in US images [61, 93]. The modified model implements a crucial DL technique termed Batch Normalization, which is used to normalize the value distribution before proceeding to the next layer [94]. Moon *et al.* proposed a CNN-based method, which ensembles VGGNet [62], ResNet [65], and DenseNet [66]. An image fusion mechanism combined with different image content representations was used [95]. In [57], the Faster region-based CNN (Faster-RCNN) [96] is used for breast lesion classification in US images. In this approach, the breast lesions were classified following BI-RADS, which means they were categorized into five classes, allowing a more specific diagnosis [57].

In the abovementioned methods, the network learns a binary classification task. However, in the domain of multi-classification learning, it has been shown that networks perform better when they are assigned multiple classes compared to binary classification. In this regard, in 2021, Behboodi *et al.* proposed a multi-class classification DL-based technique for the classification of fibroadenomas, cysts, and invasive ductal carcinomas in US images with limited data. Furthermore, the US image background is considered an additional class resulting in a 4-class classification task [30]. In [97], a novel framework for breast lesion classification was proposed by adapting the VGG16 architecture with an attention module [62]. To improve the model discrepancy between classified lesions and their labels, a combination

21

of binary cross-entropy loss and the logarithm of the hyperbolic cosine loss was used [97]. Cui *et al.* introduced FMRNet, a CNN that can adaptively fuse the information of multiple tumoral regions for breast tumor classification using US images. A modified ResNet-34 is adopted as the backbone, and two novel independent modules are implemented to extract and fuse intratumoral, peritumoral, and combined-tumoral features. Moreover, a channel attention module is employed to adaptively combine the features of the three regions [98]. Ma *et al.* proposed Fus2Net, a novel CNN containing two self-designed feature extraction modules for the benign and malignant classification of breast US lesion images. The first module has max and average pooling layers to reduce the parameters, prevent overfitting and improve the generalization. The second module introduces a scale residual unit to eliminate the influence of network depth on the performance [99]. Lastly, a Local Reference Semantic Code Network (LRSCnet) was presented for breast US images classification with few labeled data. A two-stage hierarchical encoder is introduced in LRSCnet to learn high-level semantic code that incorporates powerful high-level classification information, resulting in improved generalization ability [100].

Table 3 shows the abovementioned classification methods with the datasets used and the results obtained when testing the following evaluation metrics: *Acc*, *Sen*, *Spe*, *Pre*, AUC, *F1*, Matthews Correlation Coefficient (MCC) and Negative Predictive Value (NPV).

As with the segmentation task, breast lesion classification methods still need to be improved to be applied in clinical practice. Also, most networks were evaluated on small private datasets using different quantitative evaluation metrics, which makes hampers the performance comparison. In addition, two of the mentioned methods were introduced for multi-class classification and, therefore, can not be compared to the other approaches. For this reason, there is a need for a comparative analysis of lesion classification networks using a large dataset and the same quantitative evaluation metrics.

Moreover, as stated previously in [95], an image fusion mechanism, where different image content representations are combined to include segmentation information, was applied and the resulting image was input to the network. The fused image contained the lesion boundary, which is helpful to classify the lesion. This particular technique or the approach of testing input data variations using segmentation information has not been employed in other studies. Therefore, it is pertinent to also evaluate the networks using these input data variations, as it has been proven it can improve the classification results.

2.5 Deep Learning for Breast Lesion Segmentation and Classification in Ultrasound Images

The workflow of most computer-aided diagnosis systems is usually divided into several processes, namely preprocessing, image segmentation, and classification [38], where the segmentation step generally precedes the classification step. However, since these two tasks are related and can help learn each other, methods for simultaneous breast lesion segmentation and classification methods have been investigated.

In [8], authors applied a Mask R-CNN [101] for automatic detection, segmentation, and classification of

breast lesions in US images. Mask R-CNN is an extension of the Faster R-CNN [96] that provides automatic image segmentation by detecting the tumor bounding box and delineating a contour of the tumor area prior to lesion classification as benign or malignant [8]. In 2021, Zhang *et al.* proposed a multi-task learning (SHA-MTL) model with the implementation of soft and hard attention mechanisms for breast US image simultaneous segmentation and binary classification. This model uses DenseNet121 [66] as the encoder, which is connected to an upsampling decoder by attention-gated units with soft attention mechanism. The implementation of attention mechanisms can assist the model in focusing the lesion regions and the combination of the two image analysis tasks concurrently can reinforce each other [102]. In [103], authors also resorted to attention mechanisms and proposed a Segmentation-Based Attention network (SBANet), which is a deep CNN with segmentation-based attention information for breast lesion classification. In SBANet, U-Net [70] is employed as the segmentation network, and two ResNets [65] are used as feature extraction networks. Unlike previous DL breast classification methods that only extracted features from breast US images, the authors optimized and fused medical knowledge in the network in order to extract powerful features for classification, resulting in better performance [103]. Although several automatic breast US image analysis methods have been introduced, there is still demand for more sophisticated architectures.

In Table 4, the referred DL methods are presented, as well as the datasets used and the results of testing the evaluation metrics, such as DC, *Acc, Sen, Spe, F1* and AUC.

There have been few DL methods proposed in literature for simultaneous breast lesion segmentation and classification using US images. Therefore, there is a need for more automatic methods with acceptable performance for simultaneous breast lesion segmentation and classification using US images to help overcome the current problems of breast cancer diagnosis.

As previously mentioned, MTL has been used in other applications to improve the performance of segmentation and classification tasks as they are highly related. However, in the literature of breast lesion analysis using US images, this paradigm has not been tested yet. Therefore, it is important to test the implementation of MTL for simultaneous breast lesion segmentation and classification using US images. In addition, several variations of architectures employing MTL and using extracted features from the segmentation task for classification should be developed and compared in order to evaluate the most effective approach.

Table 2: Overview of breast lesion segmentation methods for US images.

Reference	Method	Dataset	Evaluation Metrics
Zhuang et al., 2019 [81]	RDAU-NET	Ultrasound Cases; Imaging Department of the First Affiliated Hospital of Shantou University; Dataset B (UDIAT); 1062 images	DC = 84.69%; JI = 80.67%; Acc = 97.91%; Sen = 83.19%; Spe = 99.34%; Pre = 88.58%; F1 = 84.78%; AUC = 92.27%; DCL = 15.30%;
Negi et al., 2020 [82]	WGAN-RDA-UNET	Ultrasound Cases; Imaging Department of the First Affiliated Hospital of Shantou University; Dataset B (UDIAT); 1062 images	DC = 88.41%; JI = 88.44%; Acc = 98.08%; Sen = 88.37%; Spe = 99.26%; Pre = 91.17%; F1 = 89.75%; PR-AUC = 95.42%; ROC-AUC = 89.11%; DCL = 11.59%
Shareef et al., 2020 [37]	STAN	BUSIS; 562 images	DC = 91.20%; JI = 84.70%; Sen = 91.00%; FPR = 9.30%; AER = 17.60%; AHE = 18.90px; MAE = 3.90px
		Dataset B (UDIAT); 163 images	DC = 78.20%; JI = 69.50%; Sen = 80.10%; FPR = 26.60%; AER = 46.50%; AHE = 35.50px; MAE = 9.70px
Shareef et al., 2020 [83]	ESTAN	BUSIS; 562 images	DC = 92.00%; JI = 86.00%; Sen = 91.00%; FPR = 7.00%; AER = 16.00%; AHE = 16.40px; MAE = 3.20px
		Dataset B (UDIAT); 163 images	DC = 82.00%; JI = 74.00%; Sen = 84.00%; FPR = 22.00%; AER = 38.00%; AHE = 25.50px; MAE = 7.00px
		BUSI; 780 images	DC = 78.00%; JI = 70.00%; Sen = 80.00%; FPR = 36.00%; AER = 56.00%; AHE = 34.80px; MAE = 9.90px
Zhu et al., 2020 [85]	S2P-Net	Shenzhen People's Hospital (Private Dataset); 632 images	DC = 89.67%; JI = 83.11%; Sen = 90.47%; Pre = 90.72%; ABD = 5.60px
		BUSI; 780 images	DC = 84.70%; JI = 76.39%; Sen = 85.51%; Pre = 87.62%; ABD = 11.18px
Xue et al., 2021 [32]	GG-Net	Shenzhen People's Hospital (Private Dataset); 632 images	DC = 87.10%; JI = 79.10%; Acc = 97.40%; Sen = 86.60%; Pre = 91.30%; ABD = 5.30px; HD = 16.20px
		BUSI; 780 images	DC = 82.10%; JI = 73.80%; Acc = 96.90%; Sen = 81.20%; Pre = 86.50%; ABD = 16.40px; HD = 43.90px
Xu et al., 2021 [86]	MSSA-Net	Dataset B (UDIAT); 163 images	DC = 83.78%; JI = 76.05%; Sen = 85.63%; FPR = 19.48%; AER = 33.85%
		BUSI; 780 images	DC = 80.65%; JI = 71.90%; Sen = 81.07%; FPR = 28.96%; AER = 47.90%
		Dataset 3 (Private Dataset); 320 images	DC = 95.14%; JI = 90.76%; Sen = 94.18%; FPR = 3.80%; AER = 9.62%
Iqbal et al., 2021 [87]	MDA-Net	Dataset B (UDIAT); 163 images	DC = 87.68%; JI = 78.94%; AUC = 94.38%; HD = 10.56px
		BUSIS; 562 images	DC = 91.85%; JI = 85.80%; AUC = 95.64%; HD = 7.15px
		Ultrasound Cases; 811 images	DC = 90.41%; JI = 83.79%; AUC = 94.82%; HD = 8.01px
Punn et al., 2021 [88]	RCA-IUnet	BUSIS; 562 images	DC = 93.50%; JI = 90.40%; Acc = 98.00%; Sen = 92.00%; Pre = 95.00%; HD = 4.76px; MAE = 0.19px
		BUSI; 780 images	DC = 91.30%; JI = 88.80%; Acc = 96.90%; Sen = 88.90%; Pre = 93.80%; HD = 4.81px; MAE = 2.20px
Wang et al., 2021 [80]	RF-Net	Ultrasound Cases; BUSI; 3279 images	DC = 86.91%; JI = 79,24%; Sen = 88.22%; Pre = 88.40%; HD = 22.44px; ASD = 7.68px
		Dataset B (UDIAT); 163 images	DC = 81.79%; JI = 73.09%; Sen = 90.07%; Pre = 78.61%; HD = 18.06px; ASD = 6.54px
		Radiopaedia; 504 images	DC = 87.00%; JI = 79.54%; Sen = 88.99%; Pre = 87.87%; HD = 23.66px; ASD = 6.77px

Table 3: Overview of breast lesion classification methods for US images.

Reference	Method	Dataset	Classes	Evaluation Metrics
Han *et al.*, 2017 [89]	GoogLeNet	Samsung Medical Center, Seoul, South Korea (Private Dataset); 7408 images	Benign and Malignant	Acc = 91.23%; Sen = 84.29%; Spe = 96.07%; AUC = 96.01%
Xiao *et al.*, 2018 [90]	InceptionV3	Third Affiliated Hospital of Sun Yatsen University (Private Dataset); 2058 images	Benign and Malignant	Acc = 85.13%; Sen = 77.44%; Spe = 89.06%; AUC = 91.00%; F1 = 78.00%
Byra *et al.*, 2019 [92]	VGG19 with matching layer	Private Dataset; 882 images	Benign and Malignant	Acc = 75.30%; Sen = 87.90%; Spe = 71.80%; AUC = 87.30%
		Dataset B (UDIAT); 163 images	Benign and Malignant	Acc = 84.00%; Sen = 85.10%; Spe = 83.40%; AUC = 89.30%
		OASBUD; 100 images	Benign and Malignant	Acc = 83.00%; Sen = 80.70%; Spe = 85.40%; AUC = 88.10%
Fujioka *et al.*, 2019 [94]	GoogleNet InceptionV2	Private Dataset; 947 images	Benign and Malignant	Acc = 92.50%; Sen = 95.80%; Spe = 87.50%; AUC = 91.30%
Moon *et al.*, 2020 [95]	Ensemble method with VGGNet, ResNet and DenseNet	BUSI; 780 images	Benign and Malignant	Acc = 90.77%; Sen = 96.67%; Spe = 89.00%; Pre = 72.50%; AUC = 94.89%; F1 = 82.86%
Wang *et al.*, 2021 [57]	Faster-RCNN	Private Dataset; 4294 images	BI-RADS 1	Acc = 91.14%; Sen = 85.23%; Spe = 94.37%
			BI-RADS 2	Acc = 91.50%; Sen = 84.15%; Spe = 94.41%
			BI-RADS 3	Acc = 92.40%; Sen = 87.38%; Spe = 94.56%
			BI-RADS 5	Acc = 98.10%; Sen = 97.78%; Spe = 98.13%
Behboodi *et al.*, 2021 [30]	MobileNetV2	Dataset B (UDIAT); 163 images	Invasive Ductal Carcinomas, Fibroadeno-mas, Cysts and Background	Acc = 90.00%; Sen = 50.00%; Pre = 80.00%; F1 = 62.00%
Kalafi *et al.*, 2021 [97]	VGG16 with attention module	Dataset B (UDIAT); University Malaya Medical Centre (Private Dataset); 439 images	Benign and Malignant	Acc = 93.00%; Sen = 96.00%; Spe = 90.00%; Pre = 92.00%; F1 = 94.00%; MCC = 87.00%
Cui *et al.*, 2021 [98]	FMRNet	Dataset B (UDIAT); 163 images	Benign and Malignant	Acc = 94.50%; Sen = 94.50%; Spe = 94.50%; AUC = 95.70%
		BUSI; 647 images	Benign and Malignant	Acc = 95.40%; Sen = 95.20%; Spe = 90.90%; AUC = 98.60%
Ma *et al.*, 2021 [99]	Fus2Net	Private Dataset; 100 images	Benign and Malignant	Acc = 92.00%; Sen = 95.65%; Spe = 88.89%; Pre = 88.00%; AUC = 97.00%
Zhang *et al.*, 2021 [100]	LRSCnet	Qianfushan Hospital of Shandong Province (Private Dataset); 186 images	Benign and Malignant	Acc = 97.76%; Sen = 96.29%; Spe = 93.00%; Pre = 97.74%; AUC = 95.40%; NPV = 90.90%

Table 4: Overview of simultaneous breast lesion segmentation and classification methods for US images.

Reference	Method	Dataset	Evaluation Metrics
Chiao *et al.*, 2019 [8]	Mask R-CNN	China Medical University Hospital (Private Dataset); 307 images	**Classification:** Acc = 85.00%
Zhang *et al.*, 2021 [102]	SHA-MTL	BUSI; 647 images	**Segmentation:** DC = 81.42%; Acc = 95.56%; Sen = 81.21; Spe = 97.36%; F1 = 88.55%; **Classification:** Acc = 94.12%; Sen = 89.93; Spe = 96.13%; F1 = 92.93%
Luo *et al.*, 2021 [103]	SBANet	Cancer Center of Sun Yat-sen University (Private Dataset); 1994 images	**Classification:** Acc = 90.78%; Sen = 91.18; Spe = 90.44%; F1 = 91.46%; AUC = 95.49%

Breast Ultrasound Dataset Construction

This chapter presents the datasets of breast US images found in the literature and the methodology followed to generate the dataset used in this study.

3.1 Public Datasets

Collecting a well-defined dataset is fundamental to the research on breast lesion segmentation and classification [104]. Moreover, a large dataset is required to train a sufficiently generalized model because it provides the model with a diverse set of examples to learn from. This variability helps the model learn to recognize patterns and features that are common across a wide range of images, making it more likely to perform well on diverse images. Therefore, training a network with a larger dataset with high variability can significantly improve the performance, generalization, and robustness of the model, making it more reliable for use in clinical practice, which is our ultimate goal. For this reason, several public datasets of breast US images were explored and the ones available with the highest quality and with the desired characteristics, namely ground truth segmentation masks and classification labels, were selected. This study made use of three public datasets of breast US images, which are referred to as Dataset A, B and C. The datasets were obtained from US systems with different specifications and at different times. This section presents a detailed description of each dataset, including the location of acquisition, number of samples and types of ground truth. Overall, training a network with a larger and variable dataset can significantly improve the performance and robustness of the model, making it more reliable for use in clinical practice.

3.1.1 Dataset A

Dataset A, originally termed BUSI dataset, was acquired in 2018 at the Baheya Hospital for Early Detection and Treatment of Women's Cancer (Cairo, Egypt) using the LOGIQ E9 and LOGIQ E9 Agile US systems. The dataset includes 780 images categorized into three classes: normal, benign, and malignant. Normal images are images of healthy breast tissue (without lesion), whereas benign and malignant images contain benign and malignant lesions. In this dataset, 437 images are benign, 210 are malignant, and 133

(a) Benign lesion image (b) Malignant lesion image (c) Normal image
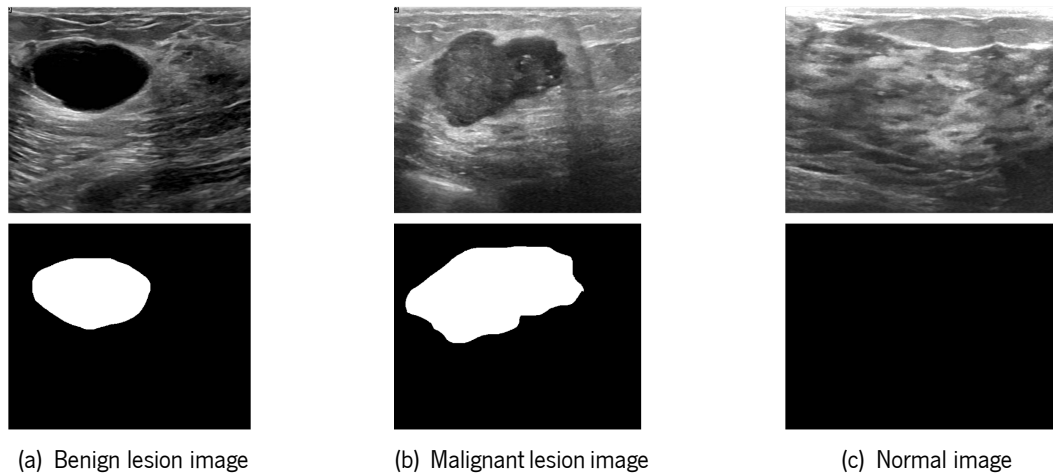
Figure 13: Examples of images in Dataset A [105]. First row represents the US image and the second row is the corresponding ground truth segmentation mask.

are normal. The images have a bit depth of 24, are stored in the PNG format with an average size of 500×500 [105].

The data were collected primarily for their application in breast cancer classification, detection, and segmentation. Therefore, the dataset provides ground truth segmentation masks for the lesions delineated by radiologists from the Baheya hospital. The binary masks have a bit depth of 1 [105].

Figure 13 displays three images from Dataset A, along with the ground truth segmentation mask below. Figure 13(a) is an example of a benign lesion with a round shape and well-defined boundaries. On the other hand, Figure 13(b) shows a malignant lesion with an angular shape. In Figure 13(c) a normal US image with the absence of lesions is presented.

3.1.2 Dataset B

Dataset B was collected in 2012 from the UDIAT Diagnostic Center of the Parc Taulí Corporation (Sabadell, Spain) using a Siemens ACUSON Sequoia C512 system. It comprises 163 images, of which 110 present benign lesions and 53 contain cancerous masses. The benign images include 65 unspecified cysts, 39 fibroadenomas, 3 lymph nodes, 2 papillomas, and 1 lymphoma. From the malignant images, 40 are invasive ductal carcinomas, 4 are ductal carcinomas *in situ*, 2 are invasive lobular carcinomas, and 7 are other unknown malignant lesions [106].

The mean image size is 760×570, the image data format is PNG. Breast lesion US images and the respective ground truth delineation, realized by experienced radiologists, are provided, along with the type of lesion information. Both original images and segmentation masks are in grayscale, presenting a bit depth of 8 [106].

Figure 14 shows examples of the aforementioned subtypes of benign lesions from Dataset B, including the delineation of the lesion. In Figure 15, examples of malignant tumors are presented, namely an

27

(a) Unspecified cyst   (b) Fibroadenoma   (c) Lymph node   (d) Papilloma   (e) Lymphoma
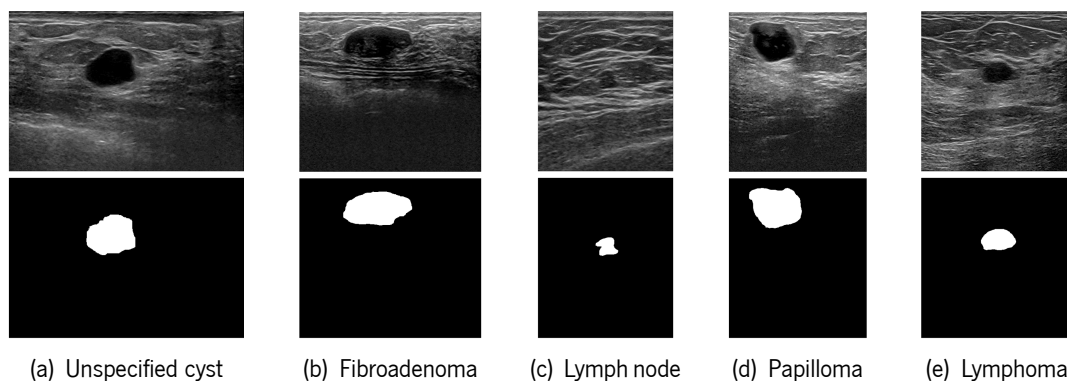
Figure 14: Examples of benign images in Dataset B [106]. First row represents the US image and the second row is the corresponding ground truth segmentation mask.

(a) Invasive ductal carcinoma   (b) Ductal carcinoma *in situ*   (c) Invasive lobular carcinoma
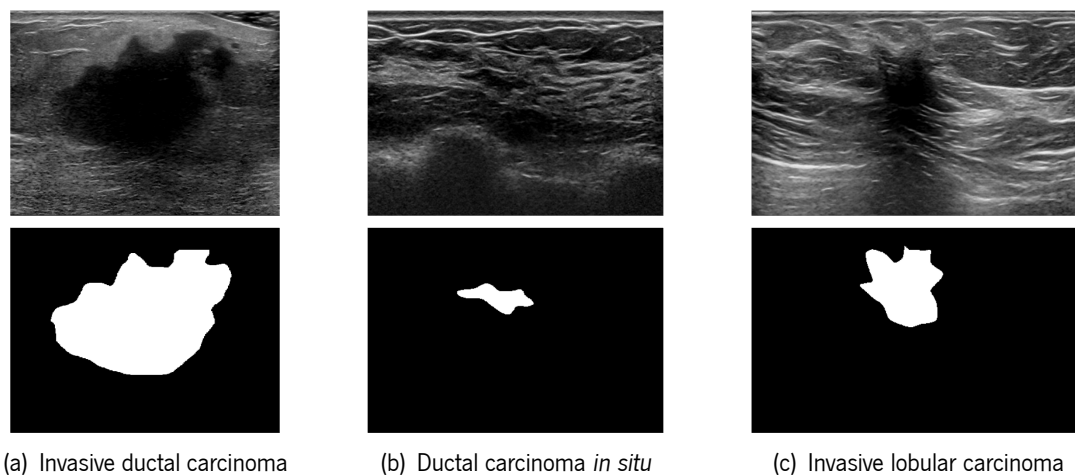
Figure 15: Examples of malignant images in Dataset B [106]. First row represents the US image and the second row is the corresponding ground truth segmentation mask.

invasive ductal carcinoma (Figure 15(a)), a ductal carcinoma *in situ* (Figure 15(b)), and an invasive lobular carcinoma (Figure 15(c)).

3.1.3   Dataset C

The third dataset, Dataset C, contains 42 US images of 128×128 pixels in the PNG format with unclassified breast lesions. It was acquired at the Imaging Department of the First Affiliated Hospital of Shantou University (Guangdong Province, China) with a GE Voluson E10 US diagnostic system. Ground truth segmentation masks created by a specialist with more than 7 years of experience from the First Affiliated Hospital of Shantou University are also available [81]. Both the US images and the binary masks are in grayscale with a bit depth of 8. Figure 16 illustrates three breast US images, and respective segmentation masks, from Dataset C.

As evident in Figures 13, 14, 15 and 16, the three datasets present differences in three aspects:

28
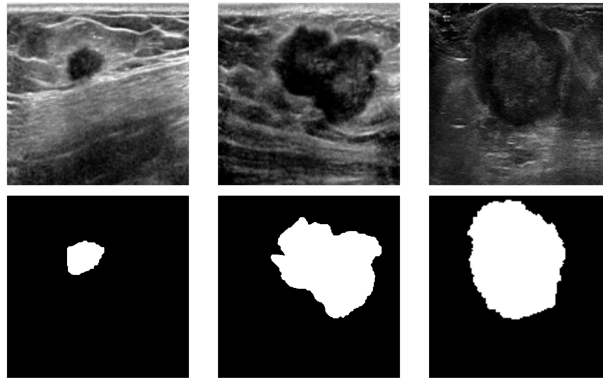
Figure 16: Examples of images in Dataset C [81]. First row represents the US image and the second row is the corresponding ground truth segmentation mask.

Table 5: Description of each dataset.

Dataset	Class	Number of Images	Average Size	Bit Depth	Ultrasound System	Notes
A	Benign Malignant Normal **Total**	437 210 133 780	500×500	24	LOGIQ E9 and LOGIQ E9 Agile	High variability in lesion size and location; Normal images available; Multiple lesions
B	Benign Malignant **Total**	110 53 163	760×570	8	Siemens ACUSON Sequoia C512	High variability in lesion size and location; Ground truth labels of lesion diagnosis available
C	**Total**	42	128×128	8	GE Voluson E10	Low variability in lesion size; Mostly centered lesions; Poor image resolution; Ground truth labels of lesion class not available

speckle noise, image quality and lesion appearance. Since Dataset A was acquired more recently and with advanced equipment, there is less speckle noise and higher quality than the other datasets. Dataset A and B present high variability in lesion location and size, unlike Dataset C, where lesions are more centered in the image.

Table 5 summarizes the datasets used in this study in terms of classes and the corresponding number of images. In total, we collected 985 breast US images with ground truth segmentation masks. Within these images, 547 contain benign lesions, 263 represent cancerous masses, 42 are unspecified breast lesions, and 133 have no lesions.

3.2 Constructed Dataset

An extensive multi-center dataset was created for this study by combining images from three sources: Datasets A, B, and C. The aim was to create a dataset with variability to allow the generation of robust and flexible DL algorithms. Since we intended to explore the segmentation of breast lesions and their
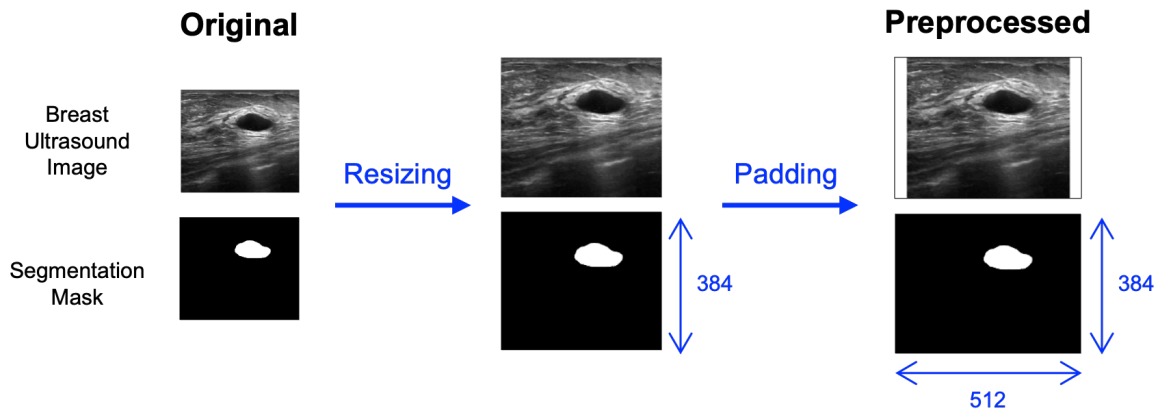
Figure 17: Summary of the preprocessing steps.

classification as benign or malignant, images categorized as normal (absence of lesions) were excluded. Thus, all samples in the constructed dataset contain one or more lesions.

Preprocessing was executed to uniformize the images from different sources (17). This procedure was performed using MATLAB (version R2019), a programming and numeric computing platform, which provides an Image Processing Toolbox with a comprehensive set of algorithms for processing, analysis, and visualization of images [107]. Since they have different dimensions, resizing was performed, followed by white padding, to maintain the width/height ratio and obtain a final image size of 512x384. These dimensions were selected based on the fact that most images were wider than taller. The same procedure was applied to the respective binary segmentation masks, but in this case, zero-padding was added. Since breast masses are generally hypoechoic, the padding on the breast US images is white to differentiate from lesions. Contrastingly, the zero-padding (black) was inserted on the segmentation masks to make it part of the background. US images were converted to grayscale, while segmentation masks were converted to logical. Finally, all images were saved in NIFTI format with the lesion class indication in the file name.

The final constructed dataset comprises 852 breast US images, where 547 contain benign lesions, 263 include malignant masses, and 42 are unclassified. These were randomly distributed and divided into groups of 682 (80%), 85 (10%), and 85 (10%) images for training, validation, and testing, respectively. The training set is a portion of a data used to fit a model for prediction of values that are known in the training set, but unknown in other data. Furthermore, the validation set is used to evaluate the models during training, while the testing set is used to compute the final performance of the models.

Comparative Analysis of Current Deep Learning Networks for Breast Lesion Segmentation in Ultrasound Images

In this chapter, several state-of-the-art DL networks for breast segmentation in US images are evaluated and compared using an extensive dataset. The work described in this chapter resulted in the following publication in the EMBC 2022 proceedings: M. R. Ferreira *et al.*, "Comparative Analysis of Current Deep Learning Networks for Breast Lesion Segmentation in Ultrasound Images", *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* 2022, pp. 3878-3881, doi: 10.1109/EMBC48229.2022.9871091. Additionally, an abstract on this work was accepted and published in the book of abstracts of the SASYR 2022: M. Ferreira, H. Torres, B. Oliveira, P. Morais, P. Novais, and J. Vilaça, "Deep Learning Networks for Lesion Segmentation in Breast Ultrasound: A Comparative Analysis," in *2nd Symposium of Applied Science for Young Researchers*, Bragança, 2022, pp. 6–8.

4.1   Introduction

Automatic lesion segmentation in breast US images may aid medical professionals in breast cancer diagnosis, the most common type of cancer among women worldwide [32]. Tumor contours can be delineated manually or automatically. However, manual delineation is time-consuming and prone to intra and inter-observer variations [108].

Although US can offer many advantages over other imaging modalities due to its non-invasive, non-ionizing, inexpensive, and real-time nature, it presents limitations, such as speckle noise, low signal-to-noise ratio, artifacts, and shadows [47]. Therefore, automatic breast lesion segmentation in a 2D US image is a challenging task, compounded by the high variability of lesions in terms of position, size, and shape [47].

Recently, DL algorithms, particularly CNNs, have revolutionized the image processing field, showing

promising results for breast US segmentation [80]. In 2015, Ronneberger *et al.* proposed the U-Net, a pixel-to-pixel, end-to-end CNN for medical image segmentation [70]. In recent years, several architectures based on U-Net have been proposed, namely the DynUNet [73]. Moreover, other encoder-decoder architectures were also explored, such as SegResNetVAE [74]. In [76], a transformer was used as the encoder of a network termed UNETR. Specifically for the breast lesion segmentation task, Wang *et al.* proposed RF-Net that improves performance by increasing the confidence of inconclusive pixels [80]. Iqbal *et al.* presented MDA-Net for lesion segmentation in breast US images [87]. Similarly, Xue *et al.* suggested a deep CNN, the GG-Net, operated with a global guidance block and lesion boundary detection modules to optimize the breast US lesion segmentation [32].

Although many breast US segmentation methods have been investigated, most were evaluated using relatively small private datasets with different quantitative metrics, which hampers the comparison of their performance. Thus, there is a need for a benchmark to compare state-of-the-art methods using a large and common dataset.

In the next sections a benchmark for breast lesion segmentation in 2D US images is presented. Seven recent DL architectures were evaluated on an extensive database composed of three publicly available datasets. By conducting comparative experiments, the strengths of the models can be analyzed, and the networks that yield better results can be identified.

4.2   Methodology

4.2.1   Deep Learning Networks

Seven neural networks were selected to study the performance of DL approaches for breast lesion segmentation. Figure 18 shows a graphical summary of this comparative study. Specifically, the following methods were evaluated: U-Net [70], DynUNet [73], SegResNetVAE [74], UNETR [76], RF-Net [80], MDA-Net [87], and GG-Net [32]. In Table 6, a brief description of each method is presented. U-Net and the models derived from it were selected to identify whether the recently developed models for breast US segmentation, namely RF-Net, MDA-Net, and GG-Net, offer advantages over general image segmentation strategies. Particularly, UNETR was studied to quantify the added value of a transformer network, a novel architecture that has shown promising results in other image segmentation tasks [76]. RF-Net, MDA-Net and GG-Net source codes were provided by their corresponding authors, while the other architectures were obtained from MONAI, a PyTorch-based, open-source framework for DL in healthcare imaging [109]. All networks were adapted to input 2D grayscale US images and output 2D binary segmentation masks. The default parameters of each network were used for testing purposes.

The architectures of the segmentation networks used in this analysis are described below:
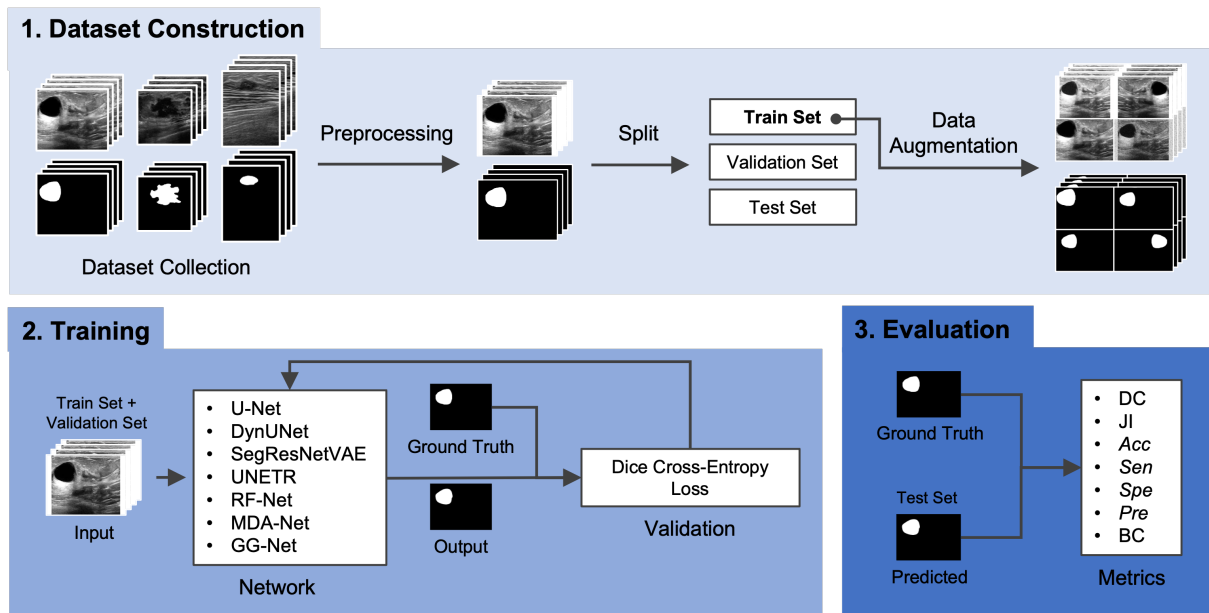
Figure 18: Graphical summary of the breast lesion segmentation comparative study.

Table 6: State-of-the-art DL segmentation networks.

Networks	Year	Description
U-Net [70]	2015	Fully convolutional encoder and decoder sub-networks with skip connections.
DynUNet [73]	2018	Dynamic U-Net with residual connection and deep supervision support.
SegResNetVAE [74]	2018	Residual U-Net with autoencoder regularization.
UNETR [76]	2021	Transformer-based network with a U-Net-like design for the encoder and decoder, and a transformer as the encoder.
RF-Net [80]	2021	Encoder-decoder-based architecture with a supervised residual representation module and a residual feedback transmission strategy.
MDA-Net [87]	2021	U-Net-inspired network with a multi-scale fusion block and a dual attention mechanism.
GG-Net [32]	2021	Deep convolutional network with a global guidance block and breast lesion boundary detection modules.

U-Net

U-Net architecture, proposed in 2015, is one of the most used approaches in biomedical image segmentation and is characterized by its symmetric U-shape (Figure 19). It has an encoder responsible for learning global contextual representations by gradually downsampling the extracted features and a decoder dedicated to pixel/voxel-wise semantic prediction by upsampling the derived representations to the input

33
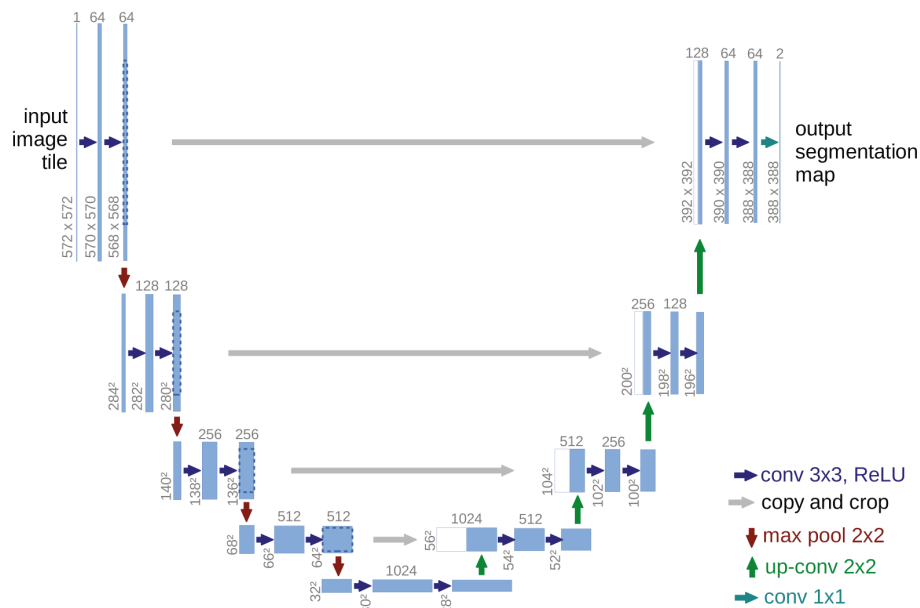
Figure 19: Schematic diagram of the U-Net architecture [70].

resolution. Additionally, skip connections combine the outputs from the encoder and decoder at multiple resolutions allowing the recovery of lost spatial information [70]. An enhanced version of U-Net, which has residual units, was used in this study.

DynUNet

DynUNet, also referred to as nnU-Net, was introduced in 2018 by Isensee *et al.* for biomedical image analysis (Figure 20). This network is a dynamic implementation of U-Net which automatically generates the decoder part for any given encoder, replaces the complex process of manual pipeline optimization with a systematic approach based on explicit and interpretable heuristic rules. This model automatically adapts to different datasets and segmentation tasks, being more flexible than U-Net since residual connection is supported in convolutional blocks, anisotropic kernel sizes and strides can be used in each layers, deep supervision can be added [73].

SegResNetVAE

SegResNetVAE is a residual U-Net with autoencoder regularization developed in 2018 (Figure 21). Here, U-Net is modified with a new architecture for encoder blocks and a variational autoencoder branch in the decoder. This branch reconstructs the input and regularizes the shared encoder in the presence of limited data, improving performance. The variational autoencoder approach allows to better group the features of the encoder endpoint [74].
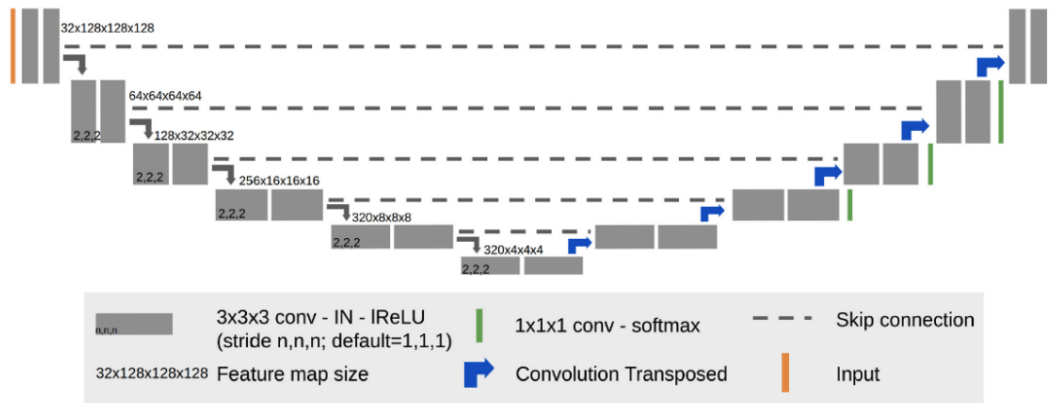
Figure 20: Schematic diagram of the DynUNet architecture for 3D image segmentation [110].
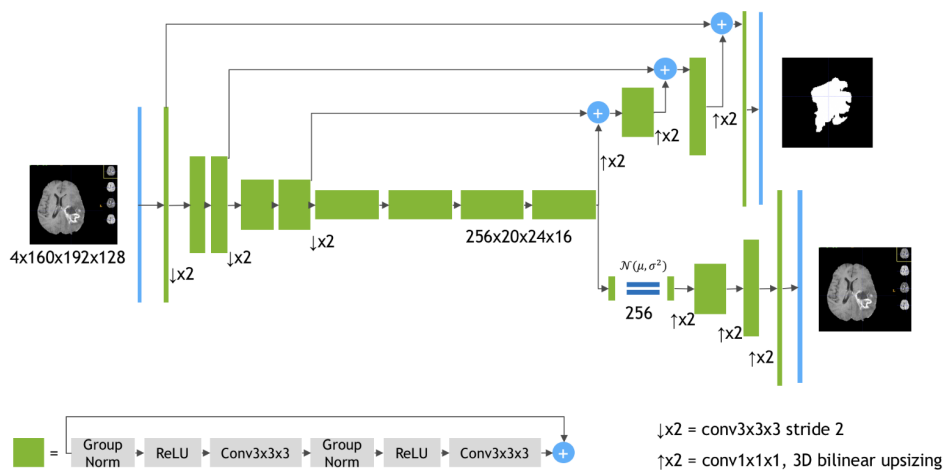
Figure 21: Schematic diagram of the SegResNetVAE architecture for 3D MRI brain tumor segmentation [74].

UNETR

In 2021, Hatamizadeh *et al.* proposed UNETR (Figure 22). The UNETR architecture is a Vision Transformer [75] generalization that uses a transformer as the encoder to learn sequence representations of the input and effectively capture the global multi-scale information, while also following a design based on U-Net for the encoder and decoder. The transformer encoder is directly connected to a decoder via skip connections at different resolutions to compute the final semantic segmentation output [76].

RF-Net

RF-Net, a network based on the encoder-decoder architecture, was introduced in 2021 by Wang *et al.* for lesion segmentation in breast US images (Figure 23). In this network, a residual representation module is employed to learn residual representations of ambiguous boundaries or confusing areas and feeds them into encoder blocks to increase the confidence of the hardly-predicted pixels, optimizing breast lesion
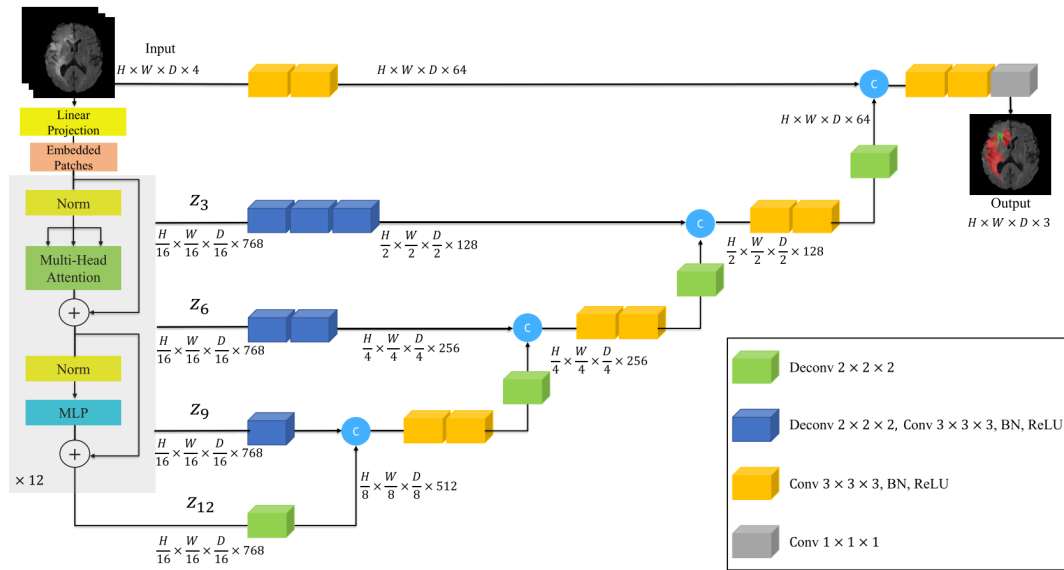
Figure 22: Schematic diagram of the UNETR architecture for 3D MRI brain tumor segmentation [76].
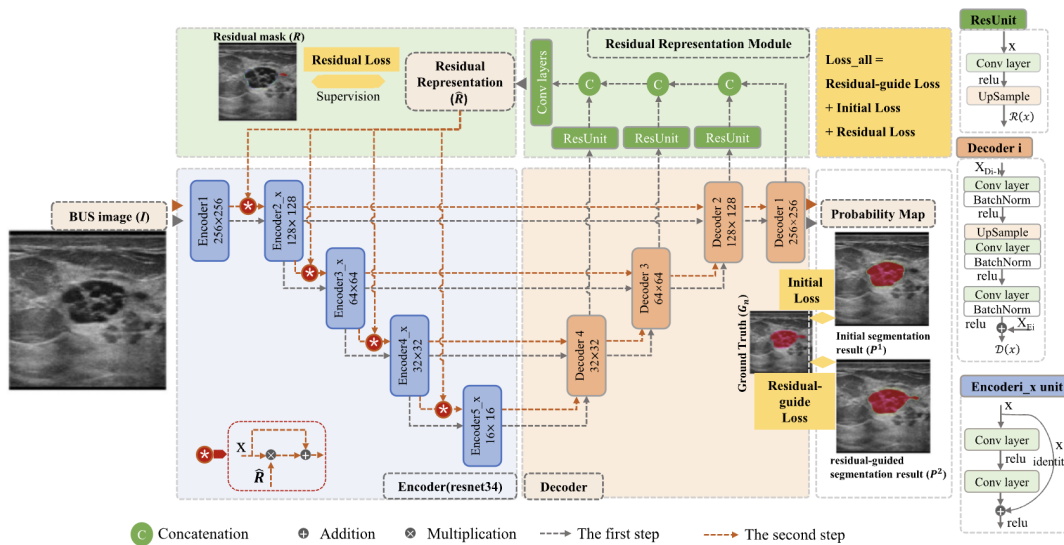
Figure 23: Schematic diagram of the RF-Net architecture [80].

segmentation. Furthermore, a residual feedback transmission strategy is designed to update the input of the encoder blocks by combining the residual representation with original features [80].

MDA-Net

In 2021, Iqbal *et al.* proposed a breast lesion segmentation network termed MDA-Net (Figure 24). It contains a multi-scale fusion block to overcome the classical fixed receptive field issues and to extract more semantic feature maps to achieve greater feature diversity. A dual-attention block is also introduced, which is a hybrid of channel-based attention and lesion attention blocks that enhances the feature representation
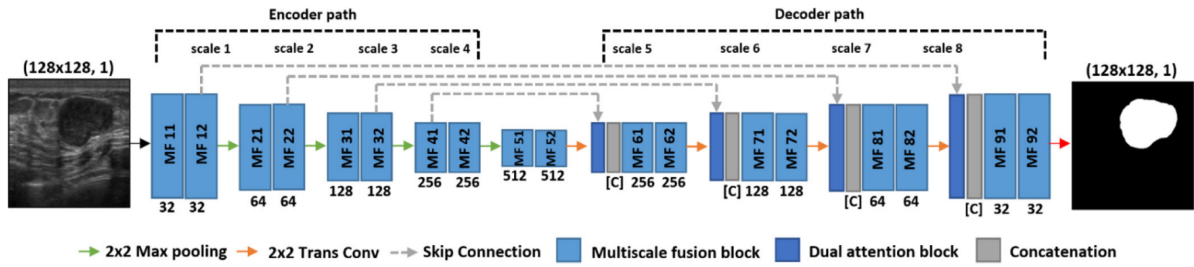
Figure 24: Schematic diagram of the MDA-Net architecture [87].
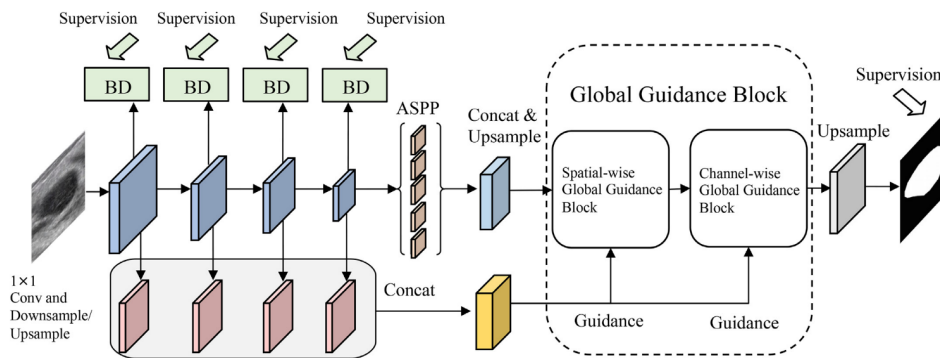
Figure 25: Schematic diagram of the GG-Net architecture [32].

capability and adaptatively learns a discriminative representation of high-level features. This combination of two attention blocks helpes to concentrate on a more relevant field of view of targets [87].

GG-Net

GG-Net was developed in 2021 by Xue *et al.* to delineate breast lesions in US images (Figure 25). It comprises a global guidance block that combines non-local features in both spatial and channel domains under the guidance of multi-layer integrated features to learn powerful contextual information. Moreover, this network integrates a breast lesion boundary detection module in shallow CNN layers to embed additional boundary maps of breast lesions for obtaining the segmentation result with high-quality boundaries [32].

4.2.2 Dataset

In this comparative analysis the constructed dataset described in Section 3.2 was used. Since the aim is to address the segmentation task, only the breast US images and respective binary segmentation masks are considered. Normalization is performed by converting the intensity values into the range of [0, 1], considering neural networks work more efficiently with these values without losing information.

4.2.3 Training Procedure

For a fair comparison, each network was trained using the Adam optimizer with a learning rate of 1e-4 and a batch size of 10 for 5000 epochs. To adjust to the tumors of varying sizes and address imbalances between the lesion regions and background, the adopted loss function was a combination of the cross-entropy and Dice loss functions.

Cross-entropy is a lower convex function that measures the differences of information content between two probability distributions. It is widely used for the classification task, and since segmentation is pixel level classification, it is appropriate for the segmentation task [111]. The Cross-Entropy loss function is defined as [112]:

$$\mathcal{L}_{CE}(y, \widehat{y}) = -\frac{1}{N} \sum_{n=1}^{N} (y_n \cdot log(\widehat{y}_n)) \tag{4.1}$$

where, y is the ground truth label, and $\widehat{y}$ is the predicted outcome the nth input US image. Dice loss function was inspired by the Dice Coefficient (DC), which calculates the similarity between two images and is the most commonly used segmentation evaluation metric [111, 112]. The Dice loss function can directly optimize the DC and is mathematically expressed as [112, 113]:

$$\mathcal{L}_{Dice}(y, \widehat{y}) = 1 - \frac{2 \sum_{n=1}^{N} (y_n \cdot \widehat{y}_n)}{\sum_{n=1}^{N} y_n^2 + \sum_{n=1}^{N} \widehat{y}_n^2} \tag{4.2}$$

The Dice Cross-Entropy loss function was used for this study, since each loss prioritizes different features in an image. Dice loss puts more emphasis on imbalanced classes and Cross-Entropy loss prioritizes the overall pixel-wise accuracy. Therefore, we combined the Dice loss with a Cross-Entropy loss to improve training stability and segmentation accuracy [114]. The Dice Cross-Entropy loss function is defined as:

$$\mathcal{L}_{DiceCE} = \lambda_{CE} \cdot \mathcal{L}_{CE} + \lambda_{Dice} \cdot \mathcal{L}_{Dice} \tag{4.3}$$

where λ_{CE} and λ_{Dice} are the weights of each loss. We set both weights to 0.5 throughout this work.

During training, the images were shuffled to prevent possible biases during data preparation. Data augmentation can achieve more complex representations of data and reduce the difference between the training set and test set, allowing CNNs to better learn the data distribution on the whole dataset [99]. Therefore, to avoid overfitting and improve performance, data augmentation techniques, including spatial and intensity-based transformations, were applied to training images. Spatial transformations included flips, zooming in, and grid distortion, whereas intensity transformations included intensity scaling and shifting, Gaussian noise and smooth addition, and contrast adjustment. Only transformations and configurations that generate images similar to real US images were considered.

Validation was executed every two epochs, and the DC was computed. To obtain the final predicted masks it was necessary to select a threshold to convert the values of the network output to values between

Table 7: Definition of the four basic cardinalities for the breast lesion segmentation task.

	True lesion	True background
Predicted lesion	TP	FP
Predicted background	FN	TN

0 and 1. Several thresholds were tested, and 0.9 was the most adequate threshold value. The model that provided the highest DC during validation and the final model were compared for each network, and the best performing model was selected.

The architectures were implemented in the Python 3.9.7 programming language using PyTorch 1.10.0, an open source ML framework, which is widely used for DL research [115]. Training was executed using NVIDIA A100 Tensor Core Graphics Processing Unit (GPU) with Ubuntu 20.04 operating system, CUDA 11.3, cuDNN 8.2, and 40GB of GPU memory.

4.2.4 Evaluation Metrics

Six widely used segmentation metrics are used to quantitatively compare the models, namely the DC, Jaccard Index (JI), accuracy (*Acc*), sensitivity (*Sen*), specificity (*Spe*), and precision (*Pre*). These indicators are mathematically expressed as follows:

$$DC = \frac{2TP}{2TP + FP + FN} \tag{4.4}$$

$$JI = \frac{TP}{TP + FP + FN} \tag{4.5}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.6}$$

$$Sen = \frac{TP}{TP + FN} \tag{4.7}$$

$$Spe = \frac{TN}{TN + FP} \tag{4.8}$$

$$Pre = \frac{TP}{TP + FP} \tag{4.9}$$

deriving from the four basic cardinalities of the confusion matrix, namely the true positives (TP), the false positives (FP), the true negatives (TN), and the false negatives (FN) [116]. These cardinalities are schematically explained in Table 7.

DC evaluates the similarity between the predicted lesions and the delineated lesions by radiologists, while JI computes the overlap percentage between the ground truth and the segmentation result [86].

Table 8: Quantitative comparison of the different segmentation methods (mean ± standard deviation %).

Networks	Evaluation Metrics						
	DC	JI	Acc	Sen	Spe	Pre	BC
U-Net	76.91 ± 24.02	67.33 ± 25.55	96.65 ± 5.40	77.36 ± 25.49	97.76 ± 5.08	81.96 ± 23.09	8.24
DynUNet	81.90 ± 23.43	74.12 ± 24.69	97.24 ± 5.04	83.50 ± 24.64	98.21 ± 4.56	83.49 ± 22.79	8.24
SegResNetVAE	81.90 ± 20.09	73.00 ± 22.05	**97.34 ± 4.37**	**84.55 ± 20.55**	98.07 ± 4.30	83.86 ± 21.06	**5.88**
UNETR	78.42 ± 24.97	69.69 ± 25.91	96.56 ± 5.69	79.39 ± 26.65	97.83 ± 5.07	81.72 ± 24.25	10.59
RF-Net	76.23 ± 25.26	66.83 ± 26.37	96.37 ± 6.07	82.59 ± 32.19	96.39 ± 6.33	63.25 ± 33.71	22.35
MDA-Net	82.03 ± 22.98	74.01 ± 23.62	97.24 ± 5.13	83.85 ± 22.68	98.39 ± 4.37	84.78 ± 22.38	7.06
GG-Net	**82.56 ± 23.31**	**74.95 ± 24.06**	97.26 ± 5.38	83.58 ± 24.21	**98.40 ± 4.61**	**85.25 ± 21.91**	7.06

Both metrics measure the same aspects and provide the same system ranking [116]. *Acc* is defined as the ratio of successfully predicted pixels to total pixels in the image [81]. *Sen*, also called True Positive Rate and recall, measures the portion of positive pixels (lesion) in the ground truth segmentation that are also identified as positive by the predicted segmentation mask being evaluated [116]. Likewise, *Spe*, also called True Negative Rate compute the proportion of negative pixels (background) in the ground truth segmentation that are also establish as negative by the predicted segmentation mask being assessed [116]. *Pre* is the ratio of the number of correctly predicted lesion pixels to the total number of predicted lesion pixels [81].

Furthermore, a second experiment was performed using a new metric, denominated "Bad Contours" (BC), where the unsuccessfully segmented cases are identified. Here, the number of results with a JI inferior to 25% was determined and used to compute the percentage of unsuccess accordingly. Overall a good segmentation performance shall have low BC but high values for all other metrics.

4.3   Results

The test images were input into the models to obtain the predicted labels. The output labels and the ground truth segmentations were then used to compute the seven metrics and assess the performance of the networks. Table 8 summarizes the mean values of the calculated metrics for each network using the complete test set. GG-Net delivered the best results, reaching a DC of 82.56%, a JI of 74.95%, a *Spe* of 98.40%, and a *Pre* of 85.25%. However, SegResNetVAE outperformed the GG-Net in terms of *Acc*, *Sen*, and BC, obtaining 97.34%, 84.55%, and 5.88%, respectively

Concerning the BC metric, we observed that the same 6 images from the test set obtained a JI lower than 25% for all networks (Figure 26). Therefore, to prevent biases in our analysis due to possible incorrectly annotated or indiscernible tumors, these 6 images were excluded. The methods' performances computed using the successfully segmented 79 images are presented in Table 9. In this case, GG-Net outperformed all networks, achieving 88.25%, 80.33%, 98.38%, 88.78%, 99.22%, and 90.07% of DC, JI, *Acc, Sen, Spe*, and *Pre*, respectively. In both tables, the results are expressed in percentages, and bold
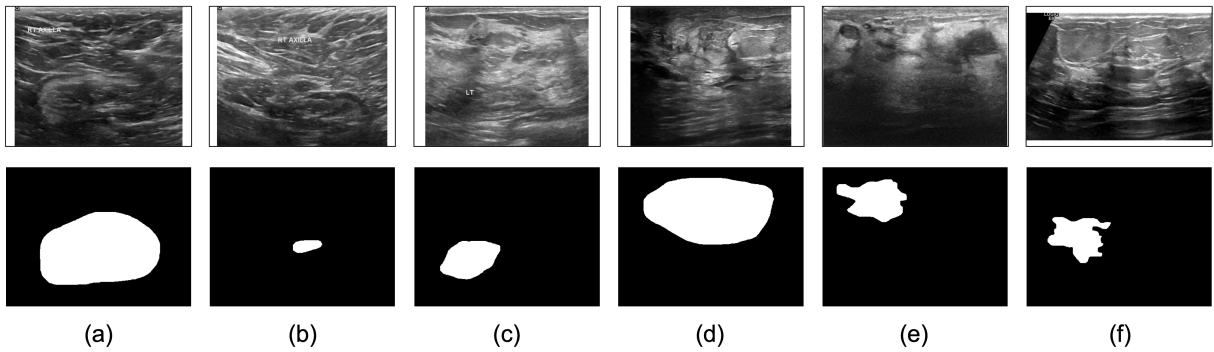
Figure 26: Images from the test set, and respective ground truth segmentation, that obtained a JI lower than 25%. (a) and (b) correspond to lymph nodes; (c) is a breast with heterogeneous echotexture; (d) represents skin thickening (possible mastitis); (e) shows two hypoechoic lesions with a posterior acoustic enhancement: on the left, an oval circumscribed nodule, and on the right, an irregular nodule with indistinct margins; (f) is a malignant lesion on fatty breast tissue.

Table 9: Quantitative comparison of the different methods, excluding images with a JI under 25% (mean ± standard deviation %).

Networks	Evaluation Metrics					
	DC	JI	Acc	Sen	Spe	Pre
U-Net	81.99 ± 15.68	72.02 ± 19.61	97.67 ± 3.49	81.35 ± 19.41	98.51 ± 3.34	86.45 ± 13.89
DynUNet	87.02 ± 13.93	79.09 ± 17.14	98.27 ± 3.03	88.11 ± 15.89	98.93 ± 2.82	87.86 ± 13.90
SegResNetVAE	86.43 ± 10.96	77.51 ± 14.91	98.20 ± 2.76	87.56 ± 14.15	98.70 ± 3.00	87.93 ± 12.54
UNETR	83.53 ± 17.02	74.51 ± 19.67	97.57 ± 4.03	83.65 ± 20.24	98.57 ± 3.44	86.13 ± 16.15
RF-Net	77.55 ± 23.83	68.13 ± 25.30	97.07 ± 4.95	88.10 ± 25.33	97.07 ± 5.29	67.40 ± 30.80
MDA-Net	87.67 ± 10.31	79.30 ± 14.02	98.30 ± 2.62	87.85 ± 14.11	99.21 ± 1.49	89.79 ± 10.08
GG-Net	**88.25 ± 10.76**	**80.33 ± 14.37**	**98.38 ± 2.97**	**88.78 ± 14.76**	**99.22 ± 2.12**	**90.07 ± 9.26**

indicates the best values for each metric.

Figure 27 illustrates examples of bad (first row), median (second row), and good (third row) segmentation outputs. The first and second columns correspond to the original US image and the delineation performed by radiologists, respectively. The columns to the right contain the predicted segmentation masks relative to different networks in this study, and each image presents the obtained DC.

4.4   Discussion

This chapter described a comparative analysis of state-of-the-art DL networks for breast lesion segmentation in US imaging. Table 8 shows that the GG-Net delivered the best performance considering DC, JI, *Spe*, and *Pre*, while SegResNetVAE achieved the highest results in terms of *Acc*, *Sen*, and BC. This means that GG-Net segmentation predictions present a greater similarity compared to the ground truth labels. On the other hand, for SegResNetVAE the ratio of successfully predicted pixels to total pixels in the image is slightly higher.
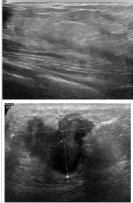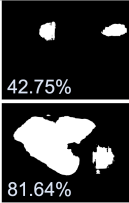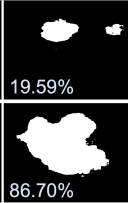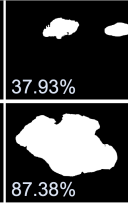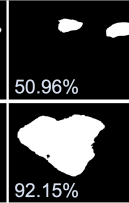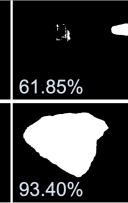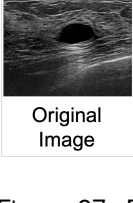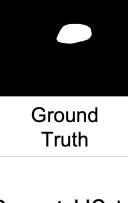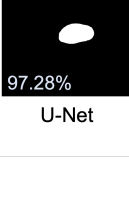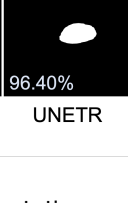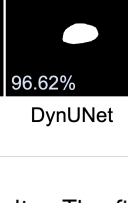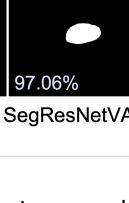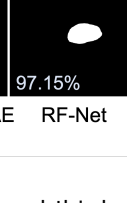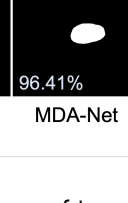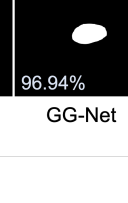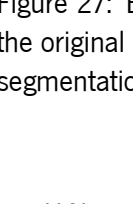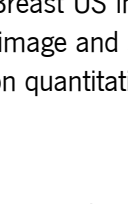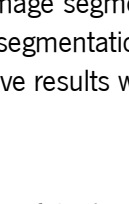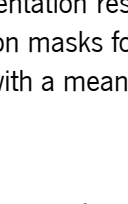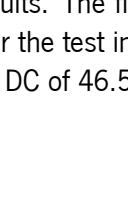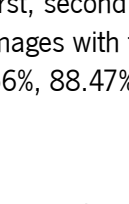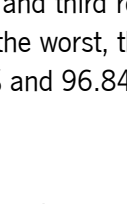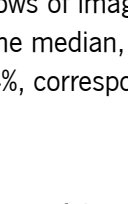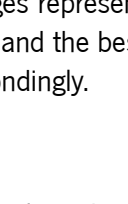
Figure 27: Breast US image segmentation results. The first, second and third rows of images represent the original image and segmentation masks for the test images with the worst, the median, and the best segmentation quantitative results with a mean DC of 46.56%, 88.47% and 96.84%, correspondingly.

U-Net presented one of the lowest performances, as expected, considering most of the evaluated networks are extensions of this architecture that significantly boost performance. Moreover, UNETR delivered inferior results compared to those shown in [76] for other imaging modalities, possibly due to the high degree of speckle noise in the breast US and tumor variability in shape, size, texture, and location. To overcome these difficulties, UNETR has to be optimized when applied to breast US. DynUNet and SegResNetVAE offered some of the highest performances, which is relevant since, like UNETR, the networks were not developed specifically for breast US. Therefore, their tuning can lead to even better results. Concerning RF-Net, it revealed lower results than anticipated, potentially related to the usage of a multi-center dataset, hampering the network's fitting to the data. Therefore, the network did not show the same robustness as it would when applied to a less variable dataset. MDA-Net achieved one of the best performances, as it implements attention mechanisms, which potentially enhanced the segmentation results. GG-Net performance met expectations since it was developed more recently and specifically for breast US image segmentation. When developing GG-Net, the authors considered that other approaches, such as U-Net, conduct convolutional operations in local regions to learn deep discriminative features, which offers unsatisfactory results. Hence, they propose the integration of all CNN layers to produce multi-level integrated features as a global guidance block to supplement more breast lesion boundary details [32]. Since the best-performing methods implement attention mechanisms, we can infer that this has a significant impact on the breast lesion segmentation task, helping locate the regions of interest with clear boundaries and accurately segment the lesion. Although GG-Net presented the best results, it requires high computational power. So if these resources are unavailable, MDA-Net or SegResNetVAE may be preferable as they require less GPU memory.

Table 9 exhibits the mean values of the metrics upon removing the 6 images with masses that showed to be difficult to segment in the abovementioned experiment. Compared to Table 8, the overall performances increased significantly, with GG-Net delivering the best performance. Moreover, when analyzing the DC metric, statistically significant differences were found between the GG-Net and three of the studied

networks, namely U-Net, UNETR, and RF-Net (p-value < 0.05 in a two-tailed paired t-test). To improve the results and to account for the variation in results on challenging images, a semi-automatic method that allows the specialist to accept or reject the outcome could be beneficial.

Figure 26 depicts the 6 excluded images in the abovementioned experiment, along with the corresponding ground truth segmentation mask. These were analyzed attentively by radiologists, who identified the possible issues that hampered lesion segmentation. Images (a) and (b), labeled benign, correspond to lymph nodes. Although these are classified as benign lesions, they are located in the axillary region, in contrast to other lesions. Since lymph nodes are a minority in the used dataset and most lesions are situated in the center of the breast, these images were more challenging to segment using the networks. In the image (c), which is originally classified as benign, a breast with heterogeneous echotexture is seen. After analysis, the radiologists stated that a lesion is not visible, so the segmentation mask is incorrect. As for image (d), the experts have determined that the lesion is correctly segmented and classified as benign, consisting of skin thickening, possibly mastitis. Since it has a different appearance compared to other lesions, its segmentation is complex. Images (e) and (f), originally classified as malignant, have incorrect ground truth segmentation masks. Image (e) shows two hypoechoic lesions, parallel to the skin, with a posterior acoustic enhancement: on the left, an oval circumscribed nodule, and on the right, an irregular nodule with indistinct margins. Regarding (f), the ground truth delineation is possibly in the wrong area, the lesion being in the center of the image.

Analyzing Figure 27, we can qualitatively evaluate the results. The first row of Figure 27 shows a small benign mass with unclear boundaries, resembling the background, which all methods failed to segment accurately. RF-Net was unable to detect a lesion, predicting an empty mask, unlike other methods that always locate a lesion. The second row shows a large malignant lesion with irregular boundaries, which is the median of the DC for all models. The third row shows a well-defined benign lesion with fewer shadows in the background, which could be why most of the evaluated methods achieved the highest DC (96.84%). Overall, the results corroborate the added value of DL strategies for the segmentation of challenging breast US images.

As future work, it is intended to go further in this study by comparing the importance of different modules, namely loss functions, optimizers, data augmentation, and adaptive learning rate strategies. Additionally, the performance comparison of DL networks for breast lesion classification is also envisioned.

4.5 Conclusion

A comparative analysis of seven state-of-the-art DL methods for the segmentation of breast lesions on 2D US imaging was presented. In summary, all networks demonstrated their potential to be used for the task of breast lesion segmentation. The GG-Net and SegResNetVAE architectures have been shown to outperform the other networks, and are therefore the best options for lesion segmentation in breast US images. The choice between these two networks depends on whether higher performance (GG-Net) or low

computational cost (SegResNetVAE) is preferred. However, additional experiments should be conducted
to improve the methods' comparison, namely studying different loss functions and hyperparameters.

Comparative Analysis of Current Deep Learning Networks for Breast Lesion Classification in Ultrasound Images

This chapter describes a comparative study of several state-of-the-art DL networks for breast lesion classification in US images using an extensive dataset. A research paper on the work described in this chapter was written and submitted to EMBC 2023.

5.1 Introduction

Accurate lesion classification as benign or malignant in breast US images is a critical task in clinical practice, requiring experienced radiologists and presenting many challenges [47]. The BI-RADS provides standardized terminology and criteria for characterizing breast lesions on US images. Radiologists use certain features on breast US images, such as shape, margin, echogenicity and orientation, to classify breast lesions according to the probability of malignancy of BI-RADS [117]. However, US presents many limitations that may hamper the diagnosis, such as the inherent speckle noise, low signal-to-noise ratio, and shadows that make the boundaries of tumors ambiguous [47]. Moreover, breast cancer diagnosis can be time-consuming and depends on the experience and attention of radiologists, being susceptible to high intra and inter-observer variability [94, 98]. Furthermore, diagnosis using breast US has a high false positive rate, which might cause unnecessary biopsies [117].

Automatic lesion classification may aid medical professionals in breast cancer diagnosis and help overcome these limitations. In this scope, CAD systems have been proposed to assist in medical image interpretation while outperforming the intra and inter-observer variability [89, 94, 98]. The breast US CAD system usually includes two main tasks: lesion segmentation and lesion classification [100].

Recently, such CAD systems using CNNs have shown great potential in medical image classification tasks due to their powerful feature learning ability [99]. In 2014, Szegedy *et al.* introduced the Inception architecture for classification and detection as GoogLeNet [61], a 22-layer deep CNN, which won the

ILSVRC 2014. GoogLeNet was later refined as InceptionV3, presenting lower error and computational cost [63]. In 2016, He *et al.* proposed the Residual Neural Network (ResNet), the top-performing entry of the ILSVRC 2015, which imports a novel architecture with a shortcut connection path that skips one or more layers, easing the training of deeper neural networks [65]. In 2017, Huang *et al.* introduced the Dense Convolutional Network (DenseNet) with dense blocks that directly connect any two layers [66]. In 2018, Sandler *et al.* proposed the MobileNetV2, a CNN based on an inverted residual structure with bottlenecking features [67]. Both DenseNet and MobileNetV2 were evaluated on various object recognition benchmark tasks, including ImageNet, and obtained significant improvements over the state-of-the-art on most of them [66, 67]. In 2019, EfficientNet was proposed along with a new compound scaling method, achieving higher accuracy and efficiency than previous CNNs [68].

Several studies have been conducted to explore the potential of these networks for breast lesion classification in US images. In 2017, Han *et al.* modified GoogLeNet to distinguish benign and malignant tumors from breast US images [89], including InceptionV3 and ResNet, for classifying masses in breast US images and concluded that InceptionV3 achieved the best performance [90]. Moon *et al.* proposed a CNN-based method, which ensembles VGGNet, ResNet, and DenseNet [95]. In 2021, Behboodi *et al.* proposed a multi-class classification approach to distinguish fibroadenomas, cysts, and invasive ductal carcinomas in breast US images using ResNet and MobileNetV2 [30].

Although several breast US classification methods have already been proposed, most studies used relatively small private datasets with different quantitative metrics, limiting the comparison between different approaches available in the literature. Also, with small center-specific datasets, it is not possible to analyze the generalization of models and quantify the added value of automatic classification in clinical practice. Therefore, there is a demand for a benchmark and a standardized evaluation method to assess and compare state-of-the-art DL models using a large common dataset.

Lesion segmentation can be performed prior to classification to narrow the Region of Interest (ROI) or just to generate ground truth delineation masks. This complementary information could potentially be used to help distinguish between benign and malignant lesions. In [95], an image fusion method that integrates the lesion shape was used to improve the performance of breast lesion classification [95]. The image fusion method consists in the creation of a fused image by combining three images, the original US image, the tumor segmentation mask, and the tumor shape image (tumor in a black background), in a single 3-channel image [95]. The authors tested several classification networks using different inputs, such as the fused image and each image that constitutes it. The results demonstrated that the fused image provides a higher performance than each image individually used as input, resulting in an Accuracy (*Acc*) and Area Under the Curve (AUC) of 90.80% and 96.73%, respectively [95]. Therefore, studying variations of input data that incorporate segmentation information may be a promising direction.

In the next sections, a benchmark for binary breast lesion classification in 2D US images is presented. Six deep neural networks were evaluated on a large database constructed from two publicly available datasets. In addition, five variations of input data to the networks were tested to study the impact of segmentation information used for classification. A comparative study was performed to analyze the
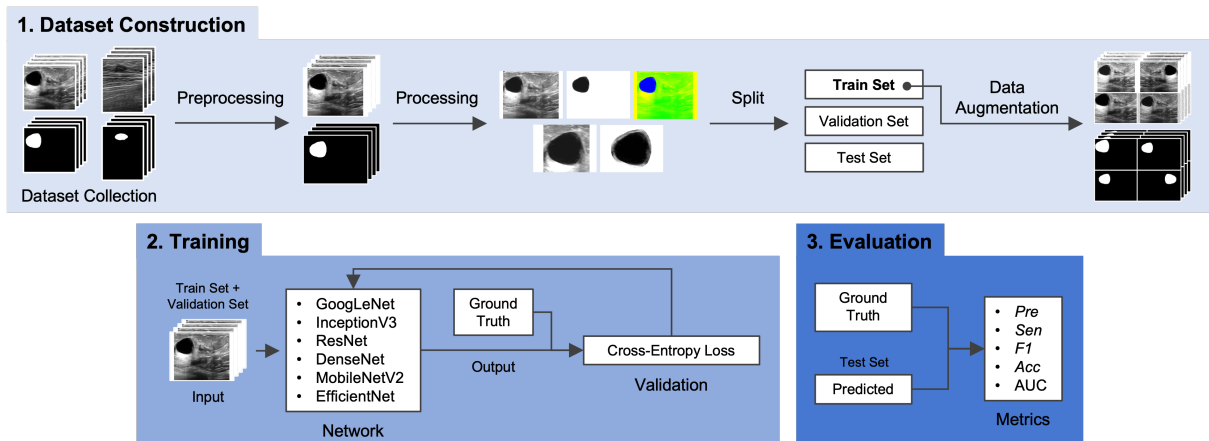
Figure 28: Graphical summary of the breast lesion classification comparative study.

strengths of each network and identify the models that yield better results.

5.2 Methodology

5.2.1 Deep Learning Networks

The performances of six relevant DL approaches for breast lesion classification were studied. Figure 28 shows a graphical summary of this comparative study.

The following networks were evaluated: GoogLeNet [61], InceptionV3 [63], ResNet [65], DenseNet [66], MobileNetV2 [67] and EfficientNet [68]. These networks were chosen because they showed high performance in object classification, particularly in distinguishing benign and malignant lesions in breast US images. GoogLeNet, InceptionV3 and MobileNetV2 architectures were obtained from PyTorch, while the others were from Project MONAI. A summary of each network is presented in Table 10.

The networks' architectures are further explained below:

GoogLeNet

In 2014, the Inception architecture was introduced by Szegedy *et al.* as GoogLeNet (Figure 29), a deep CNN with 22 layers that utilizes Inception modules [61]. The Inception modules concatenate multiple different-sized convolutional filters into a new filter reducing computational expense. This architecture's primary distinguishing feature is the better utilization of the computing resources inside the network [61].

InceptionV3

GoogLeNet was later refined as the InceptionV3 (Figure 30), containing 48 layers, to present lower error and manage memory more efficiently than previous networks [63]. The improvements InceptionV3 incorporated included factorization into smaller convolutions; batch normalization in the auxiliary classifiers;

47

Table 10: State-of-the-art DL classification networks.

Networks	Year	Description
GoogLeNet [61]	2014	22 layers deep network with Inception modules.
InceptionV3 [63]	2016	Improved version of GoogLeNet with 48 layers.
ResNet [65]	2016	Deep CNN with residual blocks, that fast-forward the activation of a layer to a deeper layer in the network.
DenseNet [66]	2017	Deep CNN with dense blocks, that connect any two layers in a feed-forward fashion.
MobileNetV2 [67]	2018	53 layers deep CNN with inverted residual blocks with bottle-necking features.
EfficientNet [68]	2020	CNN inspired by MobileNetV2, fine-tuned for obtaining high accuracy with low computational effort.
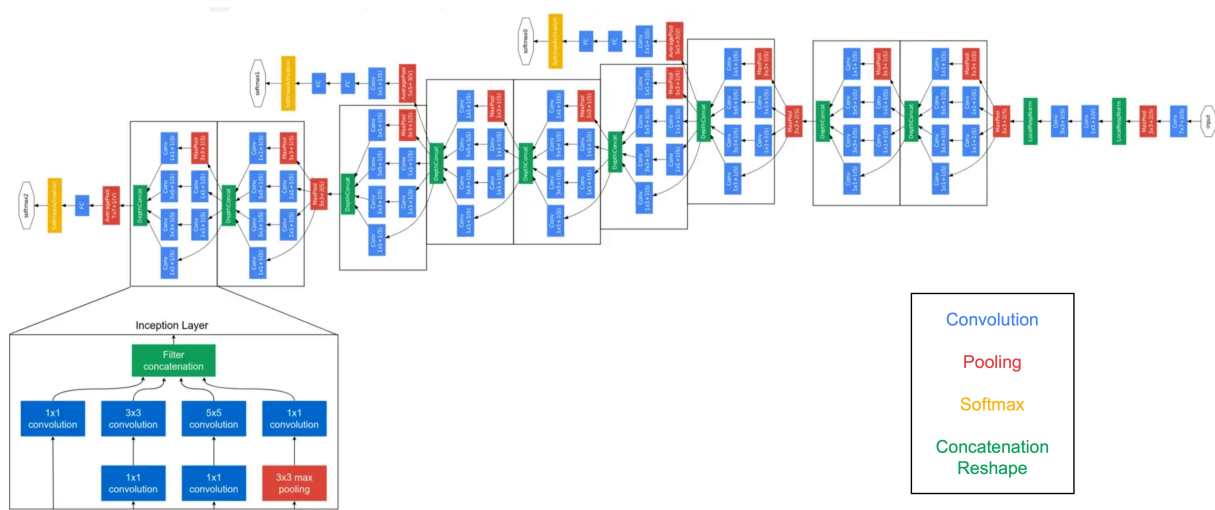
Figure 29: Schematic diagram of the GoogLeNet architecture. Adapted from [61].

and label smoothing [118]. InceptionV3 has a deeper network compared to the previous versions without compromising speed [63].

ResNet

In 2016, He *et al.* proposed the ResNet (Figure 31), equipped with residual connections between convolutional layers, that allow to skip one or more layers, to counter the vanishing gradient problem of deep CNN and allow efficient training [65]. ResNet has many variants that use the same concept but have different numbers of layers. A ResNet with 50 layers, termed ResNet50, was utilized in this work.

48
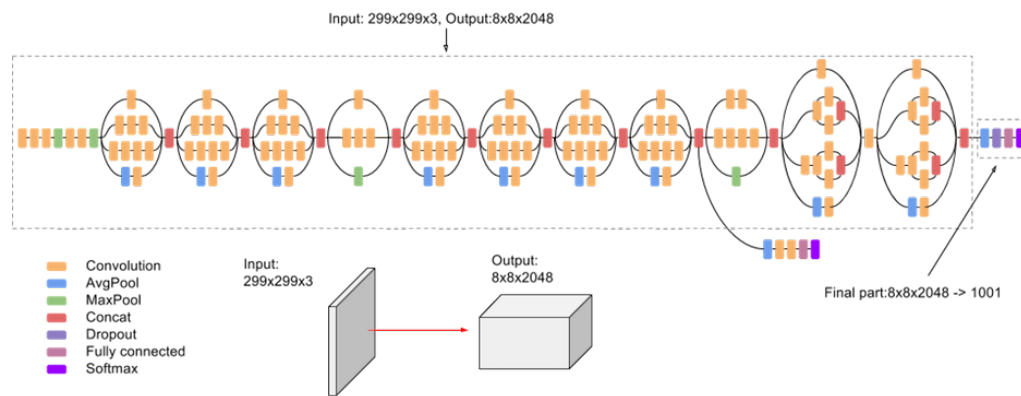
Figure 30: Schematic diagram of the InceptionV3 architecture [63].
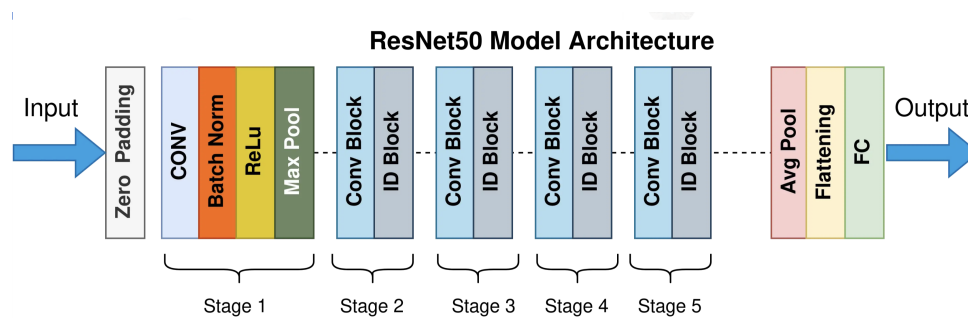
Figure 31: Schematic diagram of the ResNet50 architecture [119].

DenseNet

DenseNet, proposed by Huang *et al.* in 2017, directly connects each layer to every other layer in a feed-forward fashion within each dense block to ensure maximum information flow between layers and solve the vanishing gradient problem (Figure 32) [66]. While ResNet combines features through summation before they are passed into a layer, DenseNet combines features by concatenating them. This network improved accuracy, without any signs of performance degradation or overfitting. In this study, a DenseNet with 121 layers, called DenseNet121, was used.

MobileNetV2

In 2018, Sandler *et al.* introduced the MobileNetV2, a new mobile architecture based on an inverted residual structure where the shortcut connections are between the thin bottleneck layers, allowing faster training and better accuracy (Figure 33) [67]. This network includes an intermediate expansion layer uses lightweight depth-wise convolutions to filter features as a source of non-linearity [67].
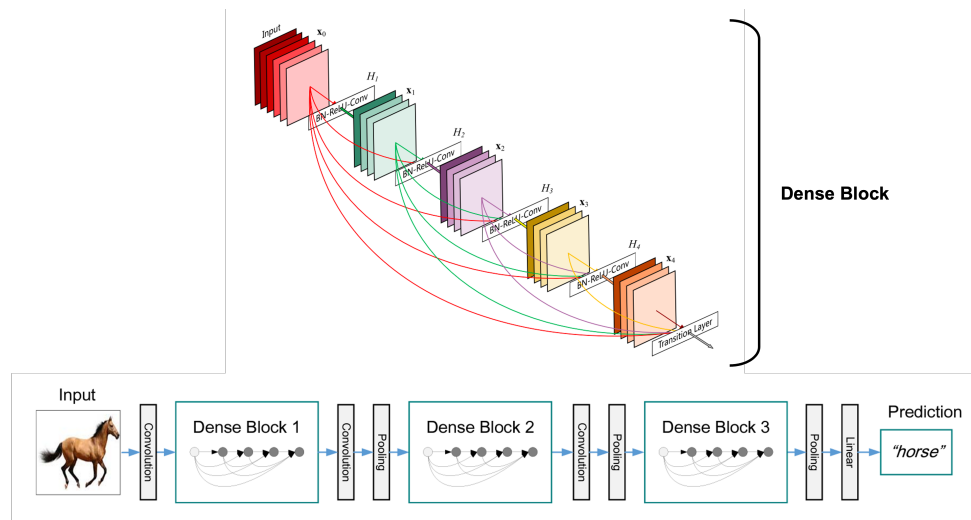
Figure 32: Schematic diagram of the DenseNet121 architecture. Adapted from [66].
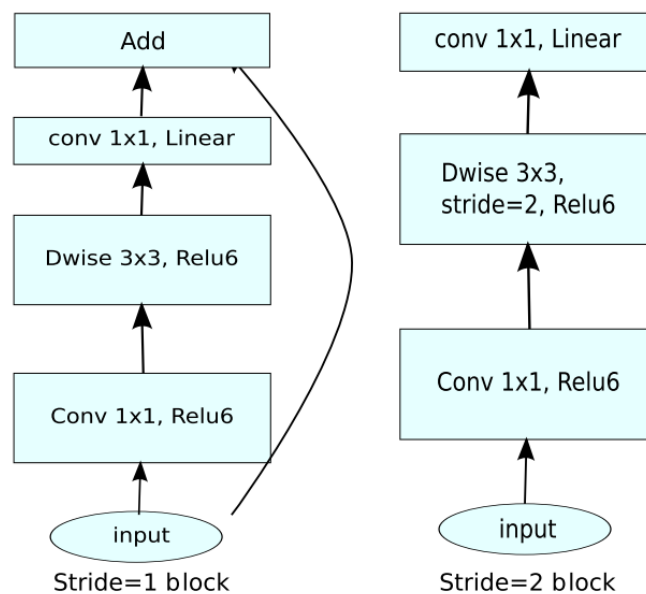
Figure 33: Schematic diagram of the MobileNetV2 architecture [67].

EfficientNet

In 2019, a new baseline network inspired by the inverted bottleneck residual blocks of MobileNetV2, termed
EfficientNet, was presented by Google AI research along with a compound coefficient to uniformly scale
network width, depth, and resolution in a principled way [68]. The EfficientNet family of models con-
sists of 8 variations from B0 to B7, where each subsequent model number means more parameters and
higher accuracy [120]. It has achieved state-of-the-art accuracy, being an order-of-magnitude smaller and
faster [68]. The baseline EfficientNet (EfficientNet-B0) was employed in this comparative study (Figure 34).

50
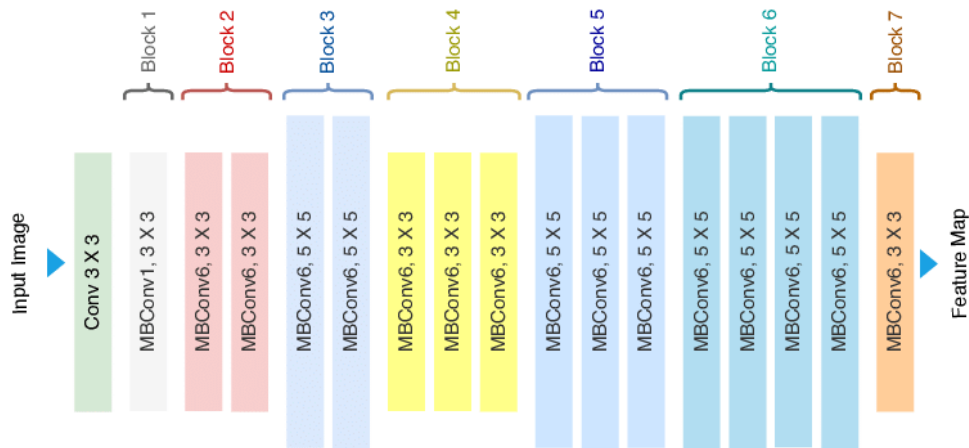
Figure 34: Schematic diagram of the EfficientNet-B0 architecture [121].

5.2.2 Dataset

The constructed multi-center dataset stated in Section 3.2 was used in this comparative analysis. However, considering the aim is to address the binary classification of breast lesions as benign or malignant, Dataset C (Subsection 3.1.3) was excluded from this study as it does not provide annotations of the lesion class. The used dataset contained 810 images with 547 benign lesions and 263 malignant tumors. The data were randomly distributed as 80% for training (648), 10% for validation (81), and 10% for testing (81).

Experiments were conducted involving the dataset, where the breast US images were reprocessed to obtain five different image variations. These experiments required the lesion segmentation masks and are explained in detail in the next section.

For training normalization is executed before training by converting the intensity values to the range between 0 and 1.

5.2.3 Input Data Variations

Using the ground truth segmentation mask, manually delineated by radiologists, four variations of the original dataset were created to add possibly useful information for breast lesion classification. The selected networks were trained and tested using the following input data variations:

(a) **Input A** - Original image;

(b) **Input B** - Enhanced lesion image. Using Input A and the respective segmentation mask, the background from the US was removed, with only the lesion remaining. The black background was replaced by a white background to better differentiate from the lesion;

(c) **Input C** - Fused image. An image fusion method inspired by [95] was applied. Input A and Input B were combined with the corresponding ground truth segmentation mask into a 3-channel image;

51
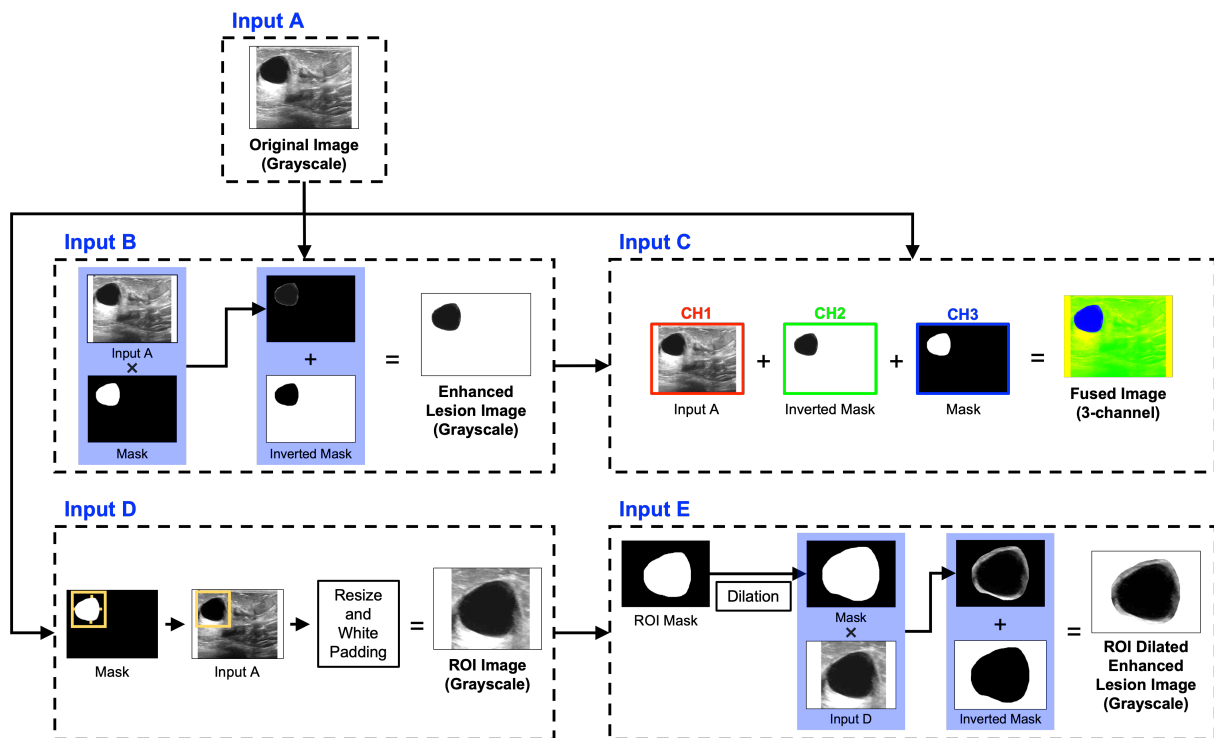
Figure 35: Examples of the image variations.

(d) **Input D** - ROI image. From the segmentation mask, the bounding box of each lesion is computed
and dilated 30 pixels from each side. Then, the image is cropped using the bounding box followed by
rescaling and padding to the original size. In specific cases, the tumor can represent a small part of
the breast US image. Thus, restricting the image to classify to the ROI instead of the original image
may improve the classification performance;

(e) **Input E** - ROI dilated enhanced lesion image. A morphological operation of dilation using a disk-shaped
structuring element is applied to the segmentation mask. Each lesion is dilated 70% of its original area,
as suggested by a radiologist. After, the pixels of the computed mask are replaced by the content of
Input D, and the background pixels are defined as 1 (white). Thus, an image corresponding to the
lesion with a small border of the background surrounding it is created. To classify a lesion, specialists
have to analyze the lesion, as well as the area around looking for characteristics, such as posterior
acoustic shadowing or a capsule [122]. For this reason, we included this input data variation in this
study.

A representation of the steps to generate these image variations is shown in Figure 35.

The datasets of Inputs A, B, and C contain 810 images with 547 benign lesions and 263 malignant
tumors. Since some images in the original dataset comprise two lesions, the datasets of Inputs D and E
are larger, containing 825 ROI images. The test sets of Inputs A, B, and C contain 55 benign images and

52

26 malignant images, while the test sets of Inputs D and E comprise 58 benign lesions and 27 malignant lesions.

5.2.4 Training Procedure

Each network was trained using the same hyperparameters for a fair comparison. Training was executed for 5000 epochs with a batch size of 10. The Adam optimizer with a learning rate of 1e-4 was used, along with the Cross-Entropy loss function (Equation 4.1).

Image shuffle was done while training to prevent possible biases related to the dataset distribution. Additionally, to prevent overfitting and improve results, training images were subjected to data augmentation techniques such as spatial and intensity-based transformations. Flips, zooming in, and grid distortion were examples of spatial transformations, whereas intensity transformations comprised intensity scaling and shifting, Gaussian noise and smooth addition, and contrast modification. Only combinations that produce images that resemble real US images were used.

Validation was performed every two epochs, and the AUC and *Acc* were computed. the Argmax function is used to convert the output probabilities of each class to the predicted class label, which is the one with the highest probability score. For each network, the model that provided the highest AUC the validation step and the final model were compared, and the one with the best results was selected.

The architectures were implemented in Python 3.9.7 and PyTorch 1.10.0 and executed using NVIDIA A100 with Ubuntu 20.04 operating system, CUDA 11.3, cuDNN 8.2, and 40GB of GPU memory.

5.2.5 Evaluation Metrics

Five evaluation metrics were used to assess the performance of the models, namely the precision (*Pre*), sensitivity (*Sen*), F1-score (*F1*), *Acc* and AUC. The mathematical formulas of *Pre*, *Sen* and *Acc* correspond to Equations 4.9, 4.7, and 4.6, respectively. *F1* and AUC are mathematically expressed as follows:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{5.1}$$

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{5.2}$$

These metrics derive from the four basic cardinalities of the confusion matrix, namely the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The meaning of the cardinalities in this context is schematically explained in Table 11.

Pre represents the proportion of positive samples that were correctly classified to the total number of positive predicted samples [123]. *Sen* is the proportion of the positive samples that were correctly classified [123]. *F1* is the harmonic mean of *Pre* and *Sen* [124]. *Acc* returns an overall measure of how much the model is correctly predicting on the entire set of data [125]. AUC reflects the ranking quality of a classifier by considering pairwise differences of scores given to samples from opposing classes [126].

Table 11: Definition of the four basic cardinalities for the breast lesion classification task.

	True benign lesion	True malignant lesion
Predicted benign lesion	TP	FP
Predicted malignant lesion	FN	TN

5.3   Experiments and Results

This section shows the results of the two conducted experiments. Firstly, six state-of-the-art DL networks
for classification were trained using input data variations and compared. Secondly, several dilation per-
centages were evaluated for the Input E, which obtained the best results in the first experiment, to identify
the most suitable percentage.

5.3.1   Comparison of Several Networks using Input Data Variations

Table 12 presents the results for the five image variations used as input to each network. Bold values
indicate the best results of that particular experiment. EfficientNet trained using Input E achieved the best
performance, obtaining a *Pre, Sen, F1, Acc,* and AUC of 97.73%, 97.65%, 97.62%, 97.65%, and 96.30%,
respectively. Receiver operating characteristic curves of each model trained using Input E are illustrated in
Figure 36, corroborating that EfficientNet obtained the highest AUC. In Figure 37, six confusion matrices
are presented as heatmaps, illustrating the performance of the different classification networks trained
using Input E on each class. The class with the most errors was the malignant class.

Figure 38 shows examples of US images from the test set. From left to right, a benign and a malignant
lesion images that all methods correctly classified; a malignant lesion that almost all methods incorrectly
classified as benign; and a benign lesion that all methods failed to classify, except when using Inputs B
and E.

5.3.2   Study of Different Lesion Border Dilation Percentages

Considering Input E provides the best results, another study was conducted where EfficientNet was trained
and tested using Input E generated with several different border percentages (0%, 30%, 50%, 70% and 90%)
to determine which percentage is the most adequate. The results of this study are presented in Table 13
and corroborate the initial suggestion from the radiologist, which was to have a border of background 70%
the size of the lesion area.

5.4   Discussion

In recent years, several DL methods have been proposed to classify benign and malignant lesions in
breast US images [30, 89, 90, 95]. In this work, we performed a comparative analysis of six networks
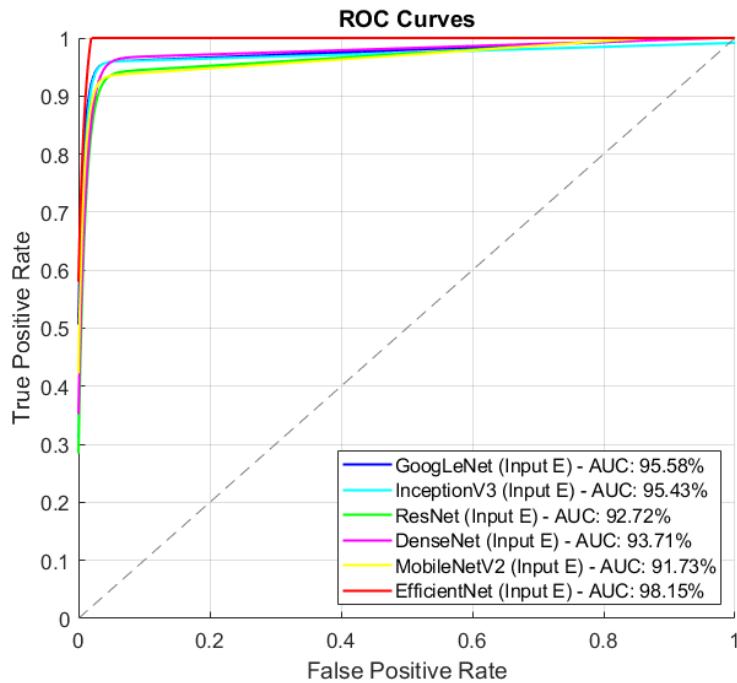
Figure 36: ROC curves of the different models trained using Input E.
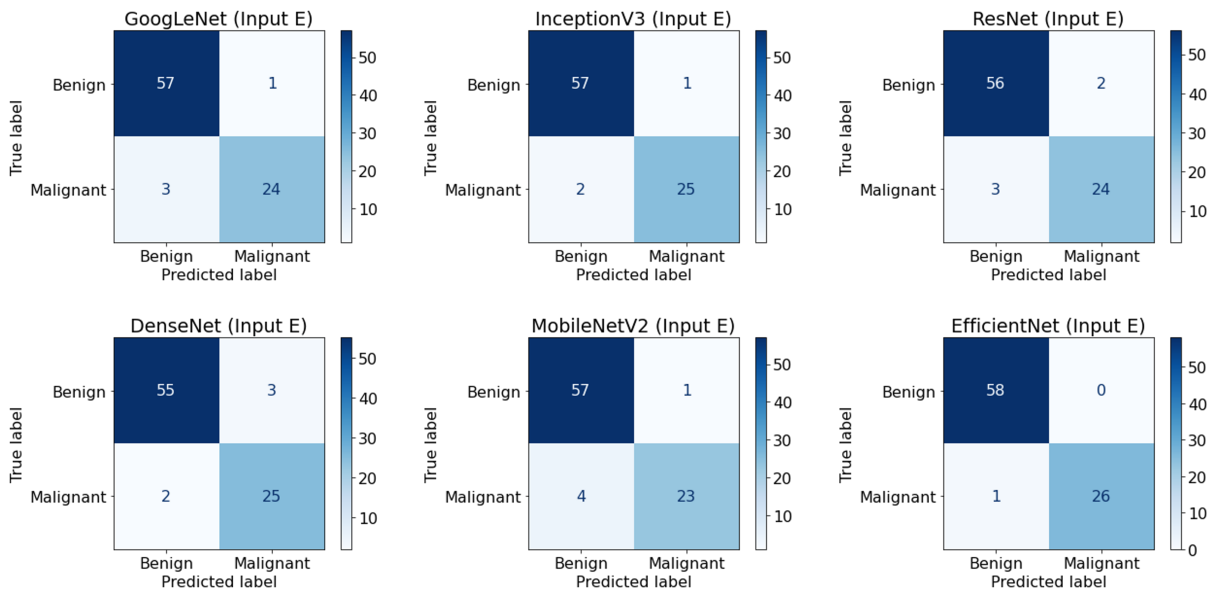
Figure 37: Confusion matrices of the different networks trained using Input E.

using five input data variations for breast lesion classification. By analyzing Table 12 and Figure 36, it is possible to determine that the overall best results were achieved by EfficientNet trained using Input E. The second-best performance was obtained by the same network trained using Input B. These input data variations maintain the overall shape of the lesion and exclude part (Input E) or the whole (Input B) background. Since the background adds entropy to the image, this process may have been beneficial to

	Ground Truth: Benign	Ground Truth: Malignant	Ground Truth: Malignant	Ground Truth: Benign
Input A (EfficientNet)	Benign	Malignant	Benign	Malignant
Input B (EfficientNet)	Benign	Malignant	Benign	Benign
Input C (EfficientNet)	Benign	Malignant	Benign	Malignant
Input D (GoogLeNet)	Benign	Malignant	Malignant	Malignant
Input E (EfficientNet)	Benign	Malignant	Benign	Benign

Figure 38: Examples of breast US images from the test set and predicted classes by the best method of each input data variation. Yellow represents the border of the lesion.

the classification task. From a clinical point of view, to classify a lesion as benign or malignant, specialists analyze the lesion, as well as the area surrounding it since it contains meaningful information regarding the lesion malignancy [122]. Therefore, we can conclude Input E allows the best classification results since it includes a border of US background around the lesion, which gives clues to identify the lesion nature.

Table 13 verifies that, for Input E, the percentage of border that provides the highest classification results is 70% of the initial lesion area.

Of all networks, EfficientNet delivered the best performance for all input data variations, except for Input D, where GoogLeNet was the most accurate at classifying the breast lesions. Being the most recent network in this study, EfficientNet is a highly efficient architecture that achieves state-of-the-art performance at a very low processing load and training and inference speeds. The main building block for EfficientNet is the mobile inverted bottleneck, originally applied in MobileNetV2. Squeeze-and-excitation optimization, Swish activation, and stochastic depth are used, which improves the performance over MobileNetV2 and other networks.

Input C, produced by an image fusion procedure, achieved better results for some of the networks compared to Input A and Input D, which lack the lesion shape, being 95.06% and 95.35%, the highest *Acc* and AUC values, respectively (EfficientNet). Although the input data variations that rely on the segmentation information, such as Inputs B, C, D, and E, demonstrated better results for most of the networks, these are only recommended if the segmentation is precise.

The confusion matrices regarding the experiment with Input E in Figure 37 provide further information about the achieved results and corroborate that EfficientNet has the lowest error. The method failed to correctly classify two malignant lesions out of 85 images, representing 2.35% of the test set. The models are better at predicting benign lesions due to class imbalance since about 32% of the images represent malignant lesions, and 68% are benign lesions. Class imbalance can affect predictive performance, specifically for the minority class (malignant lesion). A weighted loss function can be implemented to prevent

56

the impact of class imbalance on the results.

Figure 38 depicts four breast US images from the test set. The first and second columns correspond to benign and malignant lesion images that all methods correctly classified. This may be justified by the fact that the lesions have common characteristics of their classes: the benign lesion is well circumscribed, while the malignant lesion is spiculated, taller than wide, and presents posterior acoustic shadowing. The third image illustrates a small malignant lesion that was incorrectly categorized by most methods. Due to its small size, it may be difficult to assess its margin, however, it has a heterogeneous density and may be classified as non-parallel to the skin, features mainly found in malignant masses. The fourth image shows a small oval hypoechoic benign lesion that was misclassified by all networks, except when Inputs B and E were used. The background has a heterogeneous echotexture where the lesion isn't clearly highlighted, which may account for the misclassification.

In [117], the performance of a DL model was compared with the performance of radiologists in diagnosing breast cancer. The DL model achieved a high AUC of 92.4%, which was similar to the AUC obtained by two experienced radiologists and significantly higher than three inexperienced radiologists [117]. Therefore, the authors determined that DL models can improve the performance of radiologists. In the future, a similar study could be carried out to compare the performance of more models to that of radiologists.

Overall, all networks demonstrated their potential to be used for the task of breast lesion classification, most presenting *Acc* and AUC values over 90%. In future work, we intend to test different hyperparameters, optimizers, loss functions, and other adaptive learning rate strategies. Furthermore, we envision the development of a solution for simultaneous lesion segmentation and classification in breast US images.

5.5 Conclusion

A comparative analysis of six state-of-the-art DL networks for the classification of breast lesions on 2D US imaging was developed. In summary, all networks demonstrated their potential to be used for the task of breast lesion classification on US images, most presenting *Acc* and AUC values over 90%. EfficientNet outperformed the remaining networks in most experiments and obtained the highest performance when trained using Input E. In addition, this network is smaller and faster than others. However, additional tests should be conducted to improve the classification methods, namely studying a weighted loss function to overcome class imbalance problems.

Table 12: Quantitative comparison of the selected classification networks using the input data variations (mean %).

Input Data Variations	Networks	Evaluation Metrics				
		Pre	Sen	F1	Acc	AUC
Input A	GoogLeNet	86.58	86.42	86.49	86.42	84.93
	InceptionV3	90.03	90.12	90.01	90.12	87.66
	ResNet	78.38	79.01	78.35	79.01	73.39
	DenseNet	89.46	91.61	90.37	91.36	91.61
	MobileNetV2	86.11	85.19	85.43	85.19	85.03
	EfficientNet	**93.32**	**92.59**	**92.72**	**92.59**	**93.53**
Input B	GoogLeNet	96.29	96.29	96.28	96.30	95.24
	InceptionV3	95.06	95.06	95.06	95.06	94.34
	ResNet	95.72	95.06	95.14	95.06	96.36
	DenseNet	96.29	96.29	96.28	96.30	95.24
	MobileNetV2	95.06	95.06	95.06	95.06	94.34
	EfficientNet	**97.62**	**97.53**	**97.50**	**97.53**	**96.15**
Input C	GoogLeNet	87.65	87.65	87.65	87.65	85.84
	InceptionV3	87.65	87.65	87.65	87.65	85.84
	ResNet	82.04	80.25	80.70	80.25	80.38
	DenseNet	90.12	90.12	90.12	90.12	88.67
	MobileNetV2	92.59	92.59	92.59	92.59	91.50
	EfficientNet	**95.28**	**95.06**	**95.11**	**95.06**	**95.35**
Input D	**GoogLeNet**	**92.94**	**92.94**	**92.94**	**92.94**	**91.86**
	InceptionV3	89.93	89.41	89.55	89.41	89.27
	ResNet	80.87	81.18	80.97	81.18	77.30
	DenseNet	89.33	89.41	89.36	89.41	87.29
	MobileNetV2	89.32	89.41	89.23	89.41	86.30
	EfficientNet	91.76	91.76	91.63	91.76	89.02
Input E	GoogLeNet	95.32	95.29	95.24	95.29	93.58
	InceptionV3	96.47	96.47	96.45	96.47	95.43
	ResNet	94.09	94.12	94.09	94.12	92.72
	DenseNet	94.20	94.12	94.15	94.12	93.71
	MobileNetV2	94.20	94.12	94.02	94.12	91.73
	EfficientNet	**98.84**	**98.82**	**98.82**	**98.82**	**98.15**

Table 13: Quantitative comparison of EfficientNet trained using variations of Input E with different dilation percentages (mean %).

Border Percentage	Evaluation Metrics				
	Pre	Sen	F1	Acc	AUC
0%	96.53	96.47	96.49	96.47	96.42
30%	97.65	97.65	97.65	97.65	97.29
50%	97.65	97.65	97.65	97.65	97.29
70%	**98.84**	**98.82**	**98.82**	**98.82**	**98.15**
90%	97.73	97.65	97.62	97.65	96.30

<div align="right">6</div>

Development of a Deep Learning Network for Simultaneous Breast Lesion Segmentation and Classification in Ultrasound Images

This chapter describes the development of a deep neural network for simultaneous breast lesion segmentation and classification in US images using MTL. A research paper regarding the proposed network was written and prepared for submission.

6.1 Introduction

Breast tumor segmentation and classification in US images are two essential tasks for CAD of breast cancer [38]. These two tasks are highly related, and in clinical diagnosis, tumor boundary and neighborhood characteristics are useful both for an accurate segmentation and classification. However, they are challenging due to the significant shape variation of breast tumors and the poor quality of US images [48]. For this reason, various methods for performing these tasks have been explored to aid clinical practice. Chapters 4 and 5 describe comparative studies of several state-of-the-art DL networks for lesion segmentation and classification in breast US images, respectively. In fact, the studies performed allowed to comprehend different models for each task and to identify the best ones in terms of performance and computational load. Nevertheless, since these two tasks are related, training them jointly in a single network to encourage feature sharing is a promising direction to improve outcomes [79].

MTL is a learning paradigm that aims at learning multiple related tasks in parallel and improves the generalization abilities of all tasks by sharing learned representations across multiple training signals [127]. MTL have been employed in many medical image analysis tasks. In 2019, Le *et al.* proposed a MTL architecture, which combines pixel-level segmentation and global image-level classification annotations for cancer diagnosis using mammograms [78]. The proposed network is based on a FCN [69], which enables efficient feature sharing between image regions and fast prediction, with a ResNet [65] backbone [78]. The classification branch takes as input the shared representation and outputs a single class label assessing

<div align="center">60</div>

the presence or absence of cancer in the input image [78]. In 2021, Zhou *et al.* proposed a MTL method to jointly train tumor segmentation and classification in an end-to-end network for 3D automated breast US [79]. The proposed method employs a V-Net [113] backbone and consists of two sub-networks: an encoder-decoder network for segmentation and a lightweight multi-scale network for classification [79]. A classification branch was added to the bottom of the V-Net, which receives fused feature maps from different stages. It contains two FC layers and one softmax layer to predict the input volume as benign or malignant [79]. In 2022, Kang *et al.* proposed a multi-stage MTL network for thyroid nodule diagnosis, which performs three tasks in two stages: binary segmentation (background or nodule) and classification (benign or malignant) in the first stage, and multi-class segmentation in the second stage [77]. Intra- and inter-task consistent learning was employed to enforce the network learn consistent predictions for all the tasks. In the first stage, a MTL U-Net [70] with a classification branch after the encoder is used. The classification branch consists of one adaptive average pooling layer, one flatten layer, one FC layer, and one softmax layer. In the second stage, another U-Net is utilized to perform 3-class segmentation (background, benign or malignant nodule). The authors of these three studies verified that MTL boosts the performance of both tasks [77–79].

In this chapter, a MTL network for simultaneous breast lesion segmentation and classification in US images, termed MTL-Seg-Eff, is proposed to improve the outcomes of both tasks. The network architecture was based on the conclusions obtained from the studies described in Chapters 4 and 5. Thus, the proposed MTL network comprises the SegResNetVAE, for segmentation, and the EfficientNet, for classification, arranged in parallel to enhance both tasks over the single-task learning counterparts.

6.2 Methodology

In this section, the proposed MTL architecture for 2D breast US images is described. The proposed method integrates classification and segmentation in an end-to-end CNN model, which takes breast US images as inputs and produces two outputs, namely the predicted lesion segmentation mask and the predicted class. Figure 39 gives an overview of the proposed MTL network architecture for lesion segmentation and classification in breast US images.

In the comparative analysis of networks for breast lesion segmentation described in Chapter 4, we concluded that, overall, GG-Net and SegResNetVAE architectures were the best options for this task. GG-Net achieved the highest Dice Coefficient (DC), JI, *Spe* and *Pre* using the total test set, while SegResNetVAE obtained the highest *Acc*, *Sen* and BC, requiring less computational resources. For this reason, SegRes-NetVAE is used in the proposed network for segmentation. Since EfficientNet was the best-performing classification network of the comparative analysis detailed in Chapter 5, it is used for classification in the proposed network. Moreover, we use the original US image as input to the classification network.

To encourage feature sharing and optimize both networks for segmentation and classification, we created a unified architecture in which SegResNetVAE and EfficientNet are trained in parallel. As stated in

61

Figure 39: Overview architecture of the proposed network (MTL-Seg-Eff).

Chapter 4, SegResNetVAE follows an asymmetrical encoder-decoder based architecture. The encoder is larger to extract image features, and the decoder is smaller to reconstruct the segmentation mask [74]. The encoder part employs ResNet [65] blocks (green blocks in Figure 39) that consist of two convolutions with Group Normalization and ReLU, followed by an additive identity skip connection. The image dimensions are progressively down-sized by 2 using strided convolutions, and simultaneously the feature size is increased by 2. All convolutions are 3x3 with 16 initial filters, as this configuration optimizes performance compared to 8 or 24 initial filters. The input image has a size of [1,512,384], and the end of the encoder has a size of [128,64,48], which is 8 times spatially smaller than the input image. The decoder has a single ResNet block per each spatial level. In each level, the image dimensions are up-sized. Here, the number of features is reduced by a factor of 2 through 1x1 convolutions and the spatial dimension is doubled using bilinear upsampling, followed by the addition of encoder output of the equivalent spatial level. The decoder endpoint has the same size as the original image and the number of features equal to the initial filters ([16,512,384]). The decoder is followed by a 1x1 convolution which generates a 1-channel image and a Sigmoid as the activation function (Equation 6.2).

For the segmentation task we employed the Dice Cross-Entropy loss (Equation 4.3), and for the classification task, we utilized the Cross-Entropy loss (Equation 4.1). Thus, the total loss of the proposed network is the weighted summation of these two individual task losses:

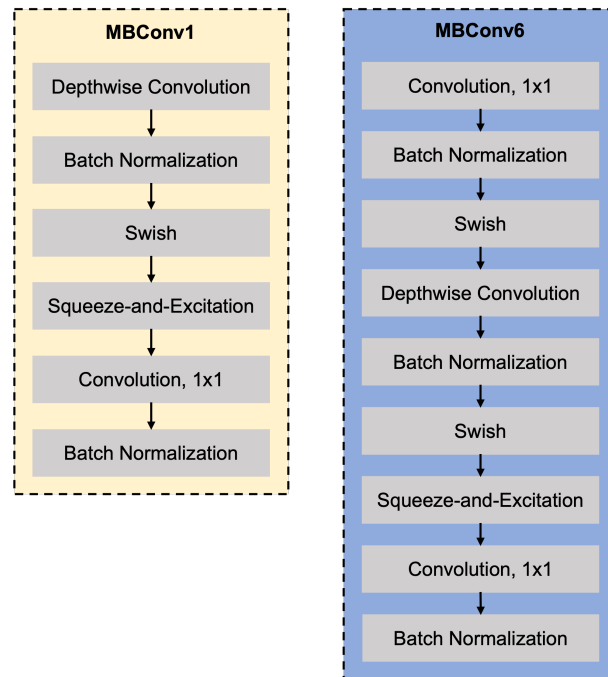$$\mathcal{L}_{Total} = \lambda_{DiceCE} \cdot \mathcal{L}_{DiceCE} + \lambda_{CE} \cdot \mathcal{L}_{CE} \tag{6.1}$$

Figure 40: Structure of MBConv1 and MBConv6 that constitute EfficientNet. Adapted from [128].

$$Sigmoid(x) = \frac{1}{1 + exp(-x))} \tag{6.2}$$

EfficientNet-B0, used for the classification task, is the baseline network of the EfficientNet group of CNN models. It is inspired by MobileNetV2, using mobile inverted bottleneck MBConv as the main building block. MBConv building blocks with 3×3 and 5×5 depthwise convolutions, squeeze-and-excitation, and swish activation. Figure 40 illustrates the structure of MBConv1 and MBConv6 layers. The EfficientNet architecture we used consists in a 3×3 MBConv1, two 3x3 MBConv6, two 5x5 MBConv6, three 3×3 MBConv1, seven 5x5 MBConv6, followed by a 3×3 MBConv6. After, one adaptive average pooling layer, one dropout layer, and a linear layer are added. LogSoftmax activation function (Equation 6.3) is used to obtain the probabilities of membership for each class.

$$LogSoftmax(x_i) = log\left(\frac{exp(x_i)}{\sum_j exp(x_j)}\right) \tag{6.3}$$

6.3 Experiments

The proposed method was evaluated and compared to several variations of its architecture. These experiments are detailed in this section.

6.3.1 Dataset and Evaluation Metrics

We evaluate our method on the dataset described in Subsection 5.2.2, consisting of 810 2D breast US images and respective ground truth segmentation tasks, where 547 are benign lesions, and 263 are malignant tumors. The data were distributed as 80% for training, 10% for validation, and 10% for testing.

To evaluate the segmentation performance we employ the following metrics: DC, precision (*Pre*), sensitivity (*Sen*), and accuracy (*Acc*). To assess classification performance, we adopt F1-Score (*F1*), accuracy (*Acc*), and area under the curve (AUC).

6.3.2 Implementation Details

The proposed method and evaluated networks for comparison were implemented in Python 3.9.7 and PyTorch 1.10.0 and trained on NVIDIA A100 40GB GPU. During training, the images were shuffled and the data augmentation techniques explained in Subsections 4.2.3 and 5.2.4 were applied. Each network was trained using the Adam optimizer with a learning rate of 1e-4 and a batch size of 10 for 5000 epochs.

The weights of each loss, λ_{DiceCE} and λ_{CE}, are set to 0.5 to contribute equally. Specifically for the Dice Cross-Entropy loss, to enhance the performance the weight of λ_{Dice} was increased, setting it to 0.7 and λ_{CE} to 0.3. SegResNetVAE and EfficientNet were trained individually, and the pre-trained weights were used to train the proposed architecture, as well as the other methods.

During validation, DC and AUC were computed to give an estimate of the model skill for the different tasks while tuning the model's hyperparameters. Considering these metrics, the models that provided the highest DC, AUC, and average of DC and AUC, along with the model at the final epoch were compared and the one, and the one with the best performance at both tasks was selected.

6.3.3 Comparison with Other Methods

To justify the performance of the proposed MTL method, we compare it with four other networks for breast lesion segmentation and classification, which are illustrated in Figure 41 and briefly explained below:

- **MTL-Seg-Eff-E** - MTL SegResNetVAE with a EfficientNet that receives Input E (Figure 35) generated from SegResNetVAE outputs as input.

- **MTL-Seg-CB** - MTL SegResNetVAE with a classification branch added after the encoder, consisting of one double convolutional layer, one adaptive average pooling layer, one flatten layer, one FC layer, and one LogSoftmax layer. The input of the classification branch is extracted features from SegResNetVAE's encoder.

- **MTL-Seg-Eff-EF** - MTL SegResNetVAE with a EfficientNet that receives features from the SegResNetVAE's encoder as input.
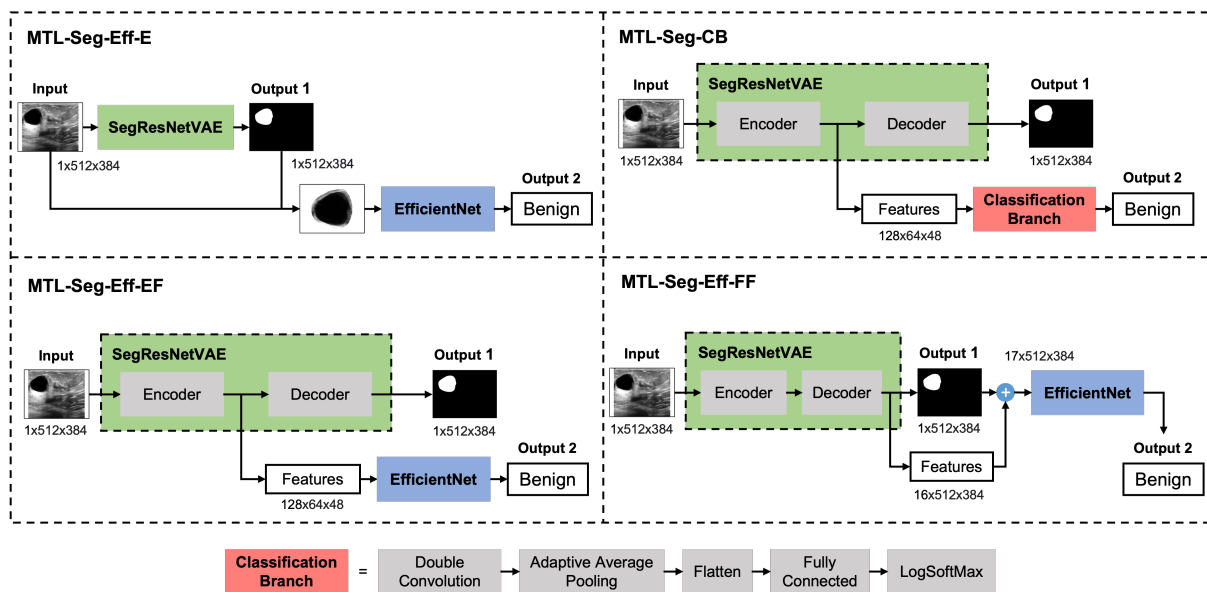
Figure 41: Simplified architectures of the compared methods.

Table 14: Quantitative comparison of the proposed method with other network variations for simultaneous segmentation (mean ± standard deviation %) and classification (mean %).

| Networks | Evaluation Metrics | | | | | | |
| | Segmentation | | | | Classification | | |
	DC	Pre	Sen	Acc	F1	Acc	AUC
SegResNetVAE	**81.47 ± 23.24**	81.49 ± 23.98	84.17 ± 23.11	**97.86 ± 2.91**	-	-	-
EfficientNet-E	-		-	-	86.82	86.42	88.98
EfficientNet	-	-	-	-	92.72	92.59	93.53
MTL-Seg-Eff-E	80.95 ± 24.19	80.35 ± 25.83	**84.43 ± 23.72**	97.79 ± 3.11	91.47	91.36	91.61
MTL-Seg-CB	80.84 ± 24.18	81.31 ± 25.83	83.20 ± 23.88	97.68 ± 3.38	88.83	88.89	86.75
MTL-Seg-Eff-EF	81.07 ± 23.89	81.64 ± 24.93	83.40 ± 24.05	97.71 ± 3.21	91.31	91.36	89.58
MTL-Seg-Eff-FF	80.84 ± 24.16	**82.00 ± 24.99**	82.76 ± 24.28	97.66 ± 3.33	91.53	91.36	92.62
MTL-Seg-Eff	80.72 ± 23.39	81.10 ± 24.55	83.35 ± 23.58	97.70 ± 3.15	**95.06**	**95.06**	**94.34**

- **MTL-Seg-Eff-FF** - MTL SegResNetVAE with a EfficientNet that receives features from after the SegResNetVAE's decoder concatenated with the original image as input.

- **MTL-Seg-Eff (Proposed Method)** - MTL SegResNetVAE with parallel EfficientNet.

6.4   Results

Table 14 gives the quantitative performance comparison of the network variations with our proposed method for simultaneous breast lesion segmentation and classification. As seen in Table 14, the proposed method outperforms other methods in all classification evaluation metrics, achieving a *F1* of 95.06%, an *Acc* of 95.06%, and an AUC of 94.34%.

In Figure 42, confusion matrices for the proposed method, as well as the other MTL variations and EfficientNet are illustrated. Additionally, Figure 43 displays the receiver operating characteristic curve of
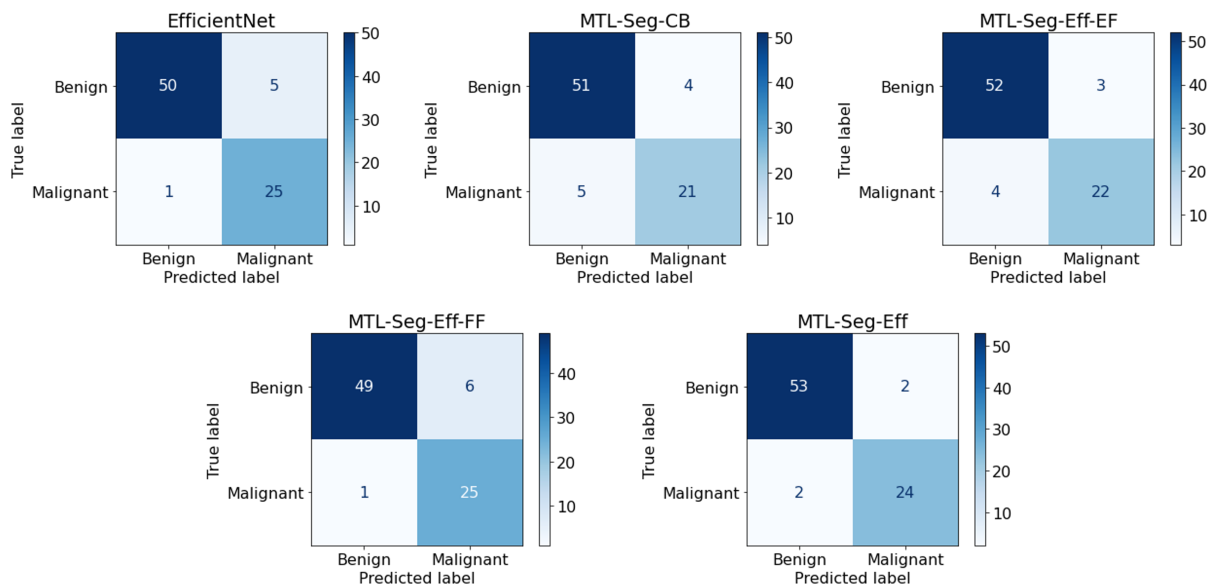
Figure 42: Confusion matrices of the EfficientNet, the proposed method and other MTL approaches.

our proposed method and other variants of the MTL approach, validating that MTL-Seg-Eff has the best performance.

The visual performance comparison of different models for segmentation and classification of breast lesions in US is shown in Figure 44.

Figure 45 displays two confusion matrices, corresponding to EfficientNet and MTL-Seg-Eff using Input E, which demonstrate that the MTL method has better classification performance than the single network. In comparison with Figure 42, it can be seen that using Input E in this case does not improve the results.

According to these experimental results, the proposed multi-task algorithm improves tumor classification performance.

6.5   Discussion

We propose a novel network architecture for simultaneous segmentation and classification of breast lesions in 2D US images that adopts MTL to optimize the performance of heterogeneous segmentation and classification tasks that share common features. Training segmentation and classification simultaneously in a unified network aims to enforce one or both of these tasks since they are highly related (for example, lesion boundary characteristics can be used as a clue both for classification and segmentation). We implemented MTL-Seg-Eff, a MTL network architecture, in which segmentation (SegResNetVAE) and classification (EfficientNet) networks are arranged in parallel and trained concurrently with the original US image as input, each delivering an output (predicted segmentation mask and classification label). To validate our proposed network, we compared it with other MTL architecture variations.
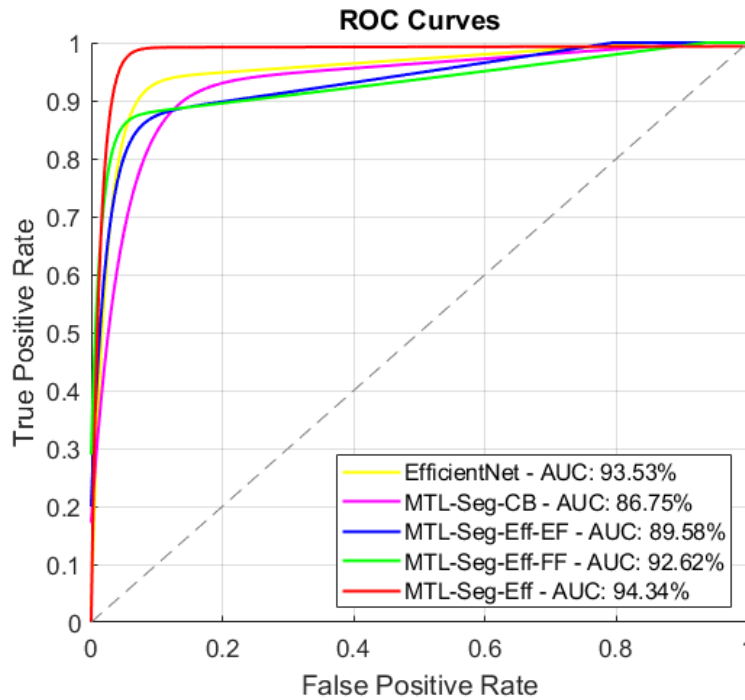
Figure 43: ROC curves of different models, including EfficientNet, MTL-Seg-CB, MTL-Seg-Eff-EF, MTL-Seg-FF, and MTL-Seg-Eff.

Table 14 gives the quantitative performance comparison of different models on breast lesion segmentation and classification. Overall, the results verify the effectiveness of our proposed MTL approach, which achieved the best classification performance, achieving a *F1* and *Acc* of 95.06%, and an AUC of 94.34%. Therefore, compared to the performance of the individual classification task (EfficientNet), the classification performance of the proposed multi-task approach has improved. To statistically corroborate this improvement, k-statistics, through Cohen's kappa coefficient, was used to determine the agreement between the classifications of the multi-task and the EfficientNet. Here, a k-value lower than 0.75 was achieved, demonstrating only moderate agreement and therefore, statistical differences between both methods. This result, together with the higher accuracy of the proposed multi-task approach, corroborates the superior performance of the proposed method in comparison with the individual classification. Additionally, we also compared these model's predictions against the ground truth classifications. EfficientNet against the ground truth obtained a k-value lower than 0.85, which means a strong agreement, while MTL-Seg-Eff achieved a k-value of 0.91, meaning an almost perfect agreement.

On the other hand, it was found that the segmentation performance did not vary significantly for each model, unlike the classification performance. In fact, no statistically significant differences were found between the proposed MTL model segmentation results and the individual segmentation model results (p-value < 0.05 in a two-tailed paired t-test). In this sense, compared with each individual task, it was found that the classification task benefits more from segmentation when using MTL since additional relevant information about the boundary and location of the lesion is given, which also allows the method to focus
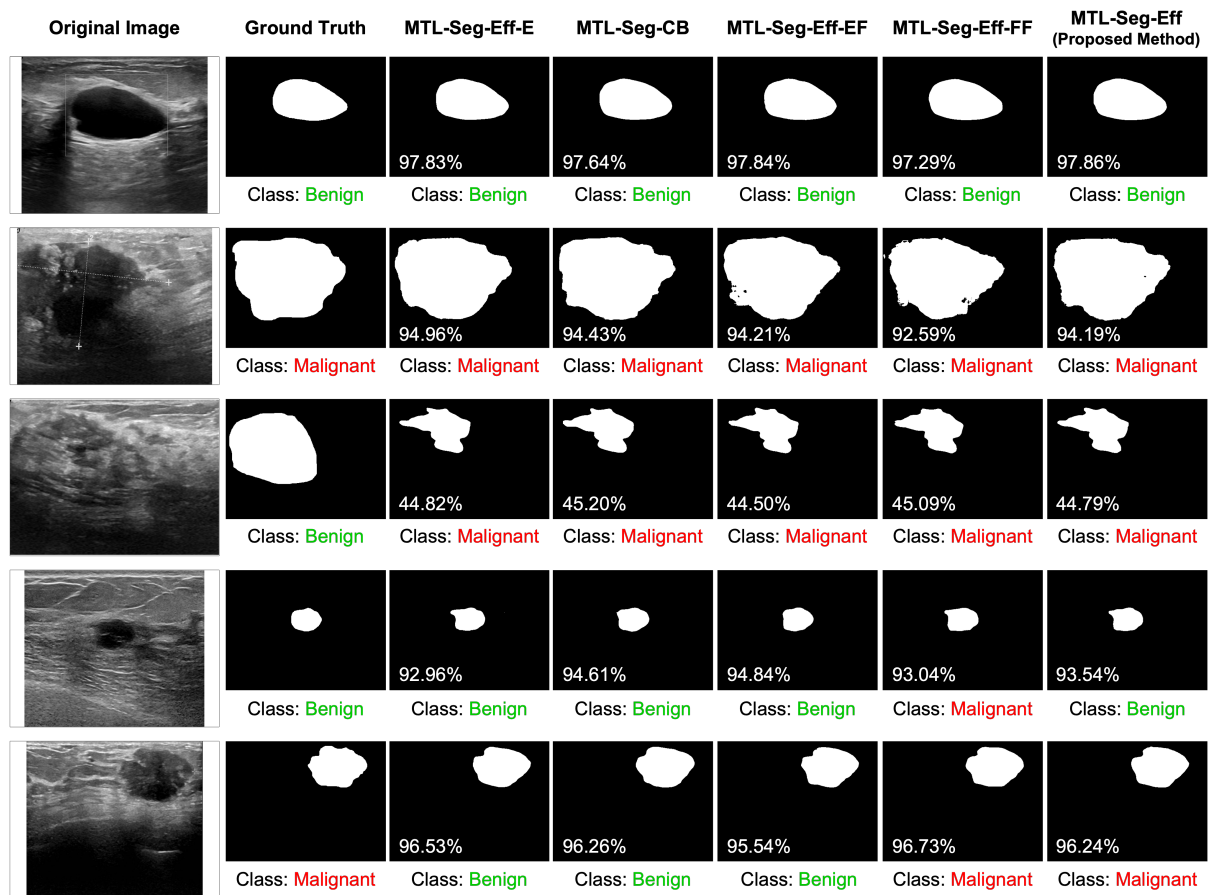
Figure 44: Visual results of different models for the segmentation and classification. From left to right: original images; ground truth for segmentation and classification; predicted labels of the MTL-Seg-Eff-E, MTL-Seg-CB, MTL-Seg-Eff-EF, MTL-Seg-Eff-FF, and MTL-Seg-Eff, respectively.
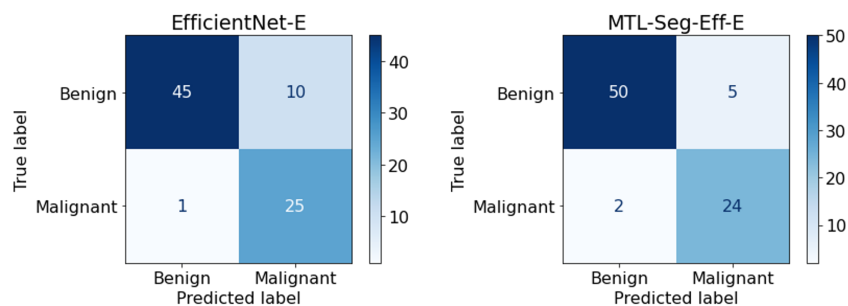

Figure 45: Confusion matrices of the EfficientNet and MTL-Seg-Eff using Input E.

on it.

Moreover, the main differences between each architecture variation are in the classification network or the input to it, allowing us to identify those that generate better results. We observed that extracting features from the SegResNetVAE's encoder and inputting the into the classification branch (MTL-Seg-CB)

68

produced the lowest classification results, which were even lower than the individual EfficientNet. MTL-Seg-Eff-EF, with a similar structure to MTL-Seg-CB, but using EfficientNet for classification, had a better overall performance than MTL-Seg-CB, particularly classification performance, proving the efficiency of EfficientNet. However, the performance was still lower than that of the individual tasks. This suggests that the original image contains more valuable information for the classification model than the extracted features from the segmentation. The use of the original US image as input to our proposed network can justify the better performance observed in our experiments. The original images can contain important information, such as texture, shape, and intensity that may be critical for accurate segmentation and classification of breast lesions. We also tested a variation of our proposed network, that incorporated features after the encoder by concatenating these features with the original image, but the overall results were still lower than the individual segmentation and classification tasks. However, the segmentation *Pre* was the highest at 82.00%. This suggests that while incorporating features after the encoder can bring some improvement, it may not be enough to fully leverage the benefits of MTL for simultaneous breast lesion segmentation and classification using US images, and that overall the proposed network is more effective.

These findings can be corroborated by Figure 42, which displays the confusion matrices of EfficientNet and the MTL architectures we developed in this study. Overall, MTL-Seg-Eff correctly classified more images, only failing 4 images from the test set (2 benign lesions and 2 malignant lesions). Furthermore, in Figure 43, the classification receiver operating characteristic curves of five different models: EfficientNet, MTL-Seg-CB, MTL-Seg-Eff-EF, MTL-Seg-Eff-FF, and MTL-Seg-Eff are depicted. The ROC curve of the proposed model is closer to the top-left corner of the graph, which represents a perfect classifier, indicating that it has a higher true positive rate and a lower false positive rate. The AUC value of the proposed model is also higher than the other models, indicating that it has a better ability to distinguish between malignant and benign lesions.

Figure 44 shows 5 examples from the test set. The first row refers to a benign lesion that all methods successfully delineated and classified, as it is smooth, oval-shaped, clearly distinct from the background, and presents posterior acoustic enhancement. The second row is a malignant lesion that all methods delineated and classified accurately since it presents an evident irregular shape and margin, is non-parallel, and presents posterior acoustic shadowing. The third image shows a benign lesion that all methods failed to classify and segment, despite being an oval and hypoechoic lesion. This is because it is a relatively small lesion surrounded by fat and fibroglandular tissue, so it blends in. In the fourth and fifth images, lesions that the proposed network accurately classified and delineated are presented. These lesions are hypoechoic and distinguishable from the surrounding background, so all methods presented a high segmentation performance. However, sometimes the margins are misplaced, causing some methods to fail. The presented examples demonstrate a clear relationship between the segmentation and classification tasks, as the performance of segmentation has a direct impact on the performance of the classification. Specifically, when the lesion is poorly segmented, the classification model is unable to accurately classify the lesion, and conversely, when the segmentation is successful, the classification is usually accurate.

In Chapter 5, we concluded that combining lesion shape information obtained from the segmentation task with the breast US image to use as input to the classification network improves performance. Specifically, using Input E, generated with the ground truth segmentation, allowed us to achieve the best results. This technique requires accurate segmentation in order to enhance the performance. Otherwise, if segmentation fails, classification will fail. Since SegResNetVAE achieved a mean DC of 81.90% in the dataset used in the comparative analysis conducted in Chapter 4, we considered that using the original breast US image as input to the classification network in MTL-Seg-Eff was a more viable option. To validate this decision, we conducted experiments to evaluated the performance of EfficientNet and MTL-Seg-Eff using Input E for classification, presented in Table 14 and Figure 45.

In these experiments we used the segmentation masks of the test set predicted by SegResNetVAE were used to generate Input E to evaluate EfficientNet-E (EfficientNet trained using Input E). However, it was found that because the segmentation DC is under 90%, there are cases where prediction might not overlap the lesion, which hampers classification performance when using the segmentation results from the SegResNetVAE instead of the segmentation ground-truth. For that reason, EfficientNet-E classification results are significantly low, compared to EfficientNet. It is worth mentioning that training SegResNetVAE and EfficientNet-E (MTL-Seg-Eff-E) in parallel instead of performing both tasks individually considerably enhances the results, from an AUC of 88.98% to 91.61%. Figure 45 corroborates that, in this experiment, adopting MTL improved classification results when using Input E. Nevertheless, in this case, using the original image as input to the network is a more advantageous approach for simultaneous breast lesion segmentation and classification using US images.

Overall, our experimental results demonstrate the added value of MTL to optimize highly related tasks. However, in order to apply this method in clinical practice, training with a larger and more varied dataset should be conducted. Therefore, in future work, the acquisition of a larger and training of the proposed model are envisioned, to improve the performance, generalization and robustness of the model, making it more reliable for use in clinical practice. Also, it is intended to adopt intra and inter-task consistent learning, a paradigm proposed in [77], to enforce the model learn consistent predictions among multiple homogeneous and heterogeneous tasks and further improve the performance of our segmentation and classification method.

6.6 Conclusion

In this chapter, we apply MTL for the joint breast lesion segmentation and classification in US images as a step towards breast diagnosis in the CAD systems.

We evaluate our proposed MTL network on a multi-center breast US image dataset. The experimental results demonstrate that the MTL architecture with parallel segmentation and classification networks improves the performance of the classification task compared with the other methods. In future work, more experiments are envisioned, such as the acquisition of a large number of breast US images with

high variability and training of the proposed model to enhance its performance, generalization and robustness, making it more reliable to be used for breast cancer diagnosis and US-guided biopsy. Moreover, other learning paradigms, such as intra and inter-task consistent learning, will be implemented in our architecture and tested.

7

Conclusions

Breast cancer has the highest incidence worldwide and the highest mortality rate among women. Early detection of breast cancer enables simpler and more efficient treatments, which improves prognosis and lowers morbidity. US imaging is one of the most used tools for breast cancer screening and diagnosis of breast cancer. Although US presents many advantages, as it is portable, radiation-free, and real-time, this process is challenging due to US's constraints, such as poor image quality and shadows. Moreover, breast US image interpretation usually involves the time-consuming task of manual delineation of the lesion and the analysis of several features, such as shape, margin, and echogenicity, which is complex due to lesion variability. Additionally, breast cancer diagnosis requires attention and experience from radiologists and is highly prone to intra and inter-observer variations. Furthermore, currently, breast US diagnosis has a high false positive rate, which might cause unnecessary biopsies. In addition to US being used for breast cancer diagnosis, it can be used for needle guidance during breast biopsy, requiring precision. The OncoNavigator project proposes a novel navigation framework to assist during US-guided breast biopsies by combining the information from the US with a collaborative medical robot and AI methods.

Considering the abovementioned problems, this dissertation had as its main objective the development of a breast lesion segmentation and classification method to help radiologists interpret US images and optimize breast cancer diagnosis. Moreover, the proposed method was developed with the intention of being incorporated in the OncoNavigator framework in order to improve the precision of US-guided breast biopsies. This dissertation focused on four tasks to achieve these goals, starting with collecting several breast US image datasets containing segmentation and classification ground truth to create a large multi-center dataset. This dataset was then used to analyze and compare several state-of-the-art DL networks for segmentation and classification of breast lesions in 2D US images, using a standardized evaluation method for a fair comparison. Finally, the knowledge obtained from the two comparative studies was applied to develop a neural network architecture for simultaneous segmentation and classification of lesions on US images.

In the comparative analysis of segmentation networks for the intended application, seven state-of-the-art DL networks were trained and tested on three aggregated datasets found in the literature. A paper on this subject was written and submitted to EMBC, a conference held by the IEEE Engineering in Medicine

and Biology Society. As for the comparative study concerning classification, six networks were trained and tested on two combined datasets. Each network was tested five times using input data variations to study the influence of adding segmentation information to images and identify the most adequate for breast lesion classification. Lastly, the best-fit networks from both studies were used to implement a network architecture adopting multi-task learning for concurrent segmentation and classification. Variations of the proposed architecture were also tested to prove its effectiveness.

The investigation work developed in the scope of this dissertation allowed us to answer the research questions presented in Chapter 1.

- **RQ1** - What are the segmentation and classification models implemented in scientific literature for breast lesions using US images?

Continuous advances have been made in research in order to apply AI for the optimization of breast cancer diagnosis. The most recent CNNs found for the segmentation of breast lesions using US images were GG-Net [32], MSSA-Net [86], MDA-Net [87], RCA-IUnet [88], and RF-Net [57]. Of these methods, the one with the lowest performance was the MSSA-Net [86] trained and tested with Dataset A, obtaining a DC of 80.65% and *Sen* of 81.07%. The best performance was achieved by RCA-IUnet [88] using the same dataset, which resulted in a DC of 91.30% and *Sen* of 92.00%. Previous methods and respective performance on different datasets are presented in Chapter 2.

The most recent models in the literature introduced or adapted for binary breast lesion classification using US images are the following: ensemble method with VGGNet, ResNet, and DenseNet [95], VGG16 with attention module [97], FMRNet [98], Fus2Net [99], and LRSCnet [100]. The ensemble method with VGGNet, ResNet and DenseNet obtained the worst performance on Dataset A with a *Sen* of 96.67%, *Acc* of 90.77% and AUC of 94.89%, while LRSCnet reached the best classification performance on a private dataset with a *Sen* of 96.29%, *Acc* of 97.76% and AUC of 95.40%. Other networks and respective performances on different datasets are presented in Chapter 2.

- **RQ2** - Which segmentation and classification models tested on a large dataset with the same evaluation metrics perform better?

The following segmentation networks were evaluated: U-Net, DynUNet, SegResNetVAE, UNETR, RF-Net, MDA-Net, and GG-Net. GG-Net delivered the best results, reaching a DC of 82.56%, a JI of 74.95%, a *Spe* of 98.40%, and a *Pre* of 85.25%. However, SegResNetVAE, lower in computational cost than GG-Net, outperformed the GG-Net in terms of *Acc*, *Sen*, and BC, obtaining 97.34%, 84.55%, and 5.88%, respectively. The performances of both networks compare to the state-of-the-art performance.

Concerning classification, GoogLeNet, InceptionV3, ResNet, DenseNet, MobileNetV2 and EfficientNet were tested. EfficientNet trained using Input E achieved the best performance, obtaining a *Pre*, *Sen*, F1, *Acc*, and AUC of 97.73%, 97.65%, 97.62%, 97.65%, and 96.30%, respectively. These results are significantly higher than the results for other networks found in the literature, which proves the effectiveness

of EfficientNet, as well as our input data variation approach that includes segmentation information and excludes irrelevant background. EfficientNet trained and tested on the original image (Input A) obtained a *Pre* of 93.32%, *Sen* of 92.59%, F1 of 92.72%, *Acc* of 92.59%, and AUC of 93.53%.

- **RQ3** - What is the influence of using different input data variations, that incorporate segmentation information, for breast lesion classification models?

When an accurate segmentation of the breast lesion is available, the integration of segmentation information, such as the boundary, may optimize classification since the lesion shape is the first feature to be analyzed when distinguishing malignant from benign lesions. If the lesion presents an irregular shape, it is considered suspicious. Furthermore, the exclusion of unnecessary background makes classification easier and improves network performance. However, the lesion must be correctly segmented. Otherwise, it is more advisable to use the original image as input for the network.

- **RQ4** - Is it possible to improve the performance of segmentation and classification tasks by using multi-task learning strategies?

The study we conducted in Chapter 6 allowed us to prove that adopting multi-task learning when performing two different tasks can improve the performance of each one of them individually. Since breast lesion classification is highly related to segmentation, multi-task learning is promising. The proposed network architecture that trains both tasks in parallel increased breast lesion classification performance while maintaining the segmentation performance.

- **RQ5** - Are the performances of the DL models adequate and comparable to human evaluation?

The DL models are adequate and comparable to human evaluation. Also, it allows a more objective and efficient evaluation, which could be used as a second opinion for medical professionals.

7.1   Limitations

During the development of this dissertation, we encountered some limitations. Firstly, we only found three datasets from different centers available in the literature that were of good quality and comparable enough to be combined, so the size of the final dataset did not meet expectations. Furthermore, only two of these datasets contained ground truth classification.

Another limitation is the fact that the images in question are isolated 2D US images of breast lesions. This makes automatic diagnosis difficult compared to manual evaluation, as radiologists have access to image sequences with different planes and settings, which allows better conditions to visualize and distinguish lesions.

Finally, there was no possibility to conduct a study with several radiologists to allow a direct comparison of the performance of the DL methods with the performance of the radiologists in order to validate the effectiveness of the models.

7.2 Future Work

Several future work can be developed within the scope of this dissertation. Firstly, regarding the dataset, we intend to partner with a hospital to acquire a more extensive dataset of US images, with a well-defined acquisition protocol and with ground truth segmentation masks and classification labels. Also, the acquisition of 2D image sequences or 3D images would be valuable, in order to have more information to segment and classify the lesions and create a more robust deep neural network architecture.

Then, it is also intended to do a study with several radiologists with different levels of experience, in order to prove the effectiveness of DL models compared to medical professionals.

Additionally, we intend to implement and test other network architectures with features that improve two simultaneous tasks, namely, the Multi-Stage Multi-Task Learning (MS-MTL) network with task consistency, proposed by Kang *et al.*, in 2022 [77]. A semi-automatic method could also be investigated.

Finally, the incorporation of the module for simultaneous segmentation and classification of breast lesions in a collaborative robot with integrated US for breast cancer procedures is envisioned.

Bibliography

[1]   P. Palhazi. "Gross Anatomy of the Breast and Axilla". In: *Breast Cancer Management for Surgeons* (2018), pp. 3–10. doi: 10.1007/978-3-319-56673-3_1 (cit. on p. 1).

[2]   I. Jatoi and M. Kaufmann. *Management of Breast Diseases*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–671. doi: 10.1007/978-3-540-69743-5 (cit. on p. 1).

[3]   S. Nawaz. "1 - The Normal Breast and Benign Diseases of the Breast". In: *Early Diagnosis and Treatment of Cancer Series: Breast Cancer*. Ed. by L. Jacobs and C. A. Finlayson. Saint Louis: W.B. Saunders, 2011, pp. 1–10. doi: https://doi.org/10.1016/B978-1-4160-4932-6.50006-8 (cit. on p. 1).

[4]   F. Ji et al. "Tumor location of the central and nipple portion is associated with impaired survival for women with breast cancer". In: *Cancer Management and Research* 11 (2019), pp. 2915–2925. doi: 10.2147/CMAR.S186205 (cit. on p. 1).

[5]   *Vetor do Stock: anatomy of the female breast side and front view. vector in cartoon simple style | Adobe Stock*. url: https://stock.adobe.com/pt/images/id/454360637?as_campaign=Freepik&as_content=api&as_audience=idp&tduid=766bd371e27968f2f01f9ba1ba7d2b49&as_channel=affiliate&as_campclass=redirect&as_source=arvato&continue-checkout=1&asset_id=454360637 (visited on 02/03/2022) (cit. on pp. 2, 3).

[6]   R. W. Ruddon. *Cancer biology*. 4th ed. New York: Oxford University Press, 2007 (cit. on p. 2).

[7]   Y. Feng et al. "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis". In: *Genes and Diseases* 5.2 (2018), pp. 77–106. doi: 10.1016/j.gendis.2018.05.001 (cit. on pp. 2, 5).

[8]   J. Y. Chiao et al. "Detection and classification the breast tumors using mask R-CNN on sonograms". In: *Medicine (United States)* 98.19 (2019), pp. 1–5. doi: 10.1097/MD.0000000000015200 (cit. on pp. 2, 6, 22, 23, 25).

[9]   Z. Klimonda et al. "Breast-lesions characterization using Quantitative Ultrasound features of peri-tumoral tissue". In: *Scientific Reports* 9.1 (2019), pp. 1–9. doi: 10.1038/s41598-019-44376-z (cit. on p. 2).

[10]   J. R. C. Sainsbury. "ABC of breast diseases: Breast cancer". In: *BMJ* 321.7263 (Sept. 2000), pp. 745–750. doi: 10.1136/bmj.321.7263.745 (cit. on p. 2).

[11]   H. Sung et al. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". In: *CA: A Cancer Journal for Clinicians* 71.3 (2021), pp. 209–249. doi: 10.3322/caac.21660 (cit. on pp. 3, 4, 12).

[12]   C. G. Yedjou et al. "Health and Racial Disparity in Breast Cancer". In: *Advances in Experimental Medicine and Biology* 1152 (2019), pp. 31–49. doi: 10.1007/978-3-030-20301-6_3 (cit. on p. 4).

[13]   Z. Momenimovahed and H. Salehiniya. "Epidemiological characteristics of and risk factors for breast cancer in the world". In: *Breast Cancer: Targets and Therapy* 11 (2019), pp. 151–164. doi: 10.2147/BCTT.S176070 (cit. on p. 4).

[14]   G. Natale et al. "The lymphatic system in breast cancer: Anatomical and molecular approaches". In: *Medicina (Lithuania)* 57.11 (2021). issn: 16489144. doi: 10.3390/medicina57111272 (cit. on p. 4).

[15]   R. W. Carlson et al. "Invasive Breast Cancer". In: *Journal of the National Comprehensive Cancer Network* 9.2 (2011), pp. 136–222 (cit. on p. 4).

[16]   A. Stachs et al. "Benign breast disease in women". In: *Deutsches Arzteblatt International* 116.33-34 (2019), pp. 565–573. doi: 10.3238/arztebl.2019.0565 (cit. on p. 4).

[17]   J. M. Dixon. *ABC Breast Diseases*. Vol. 4. 0. 2012, p. 168 (cit. on p. 4).

[18]   E. Shim et al. "Lymphoma affecting the breast: A pictorial review of multimodal imaging findings". In: *Journal of Breast Cancer* 16.3 (2013), pp. 254–265. doi: 10.4048/jbc.2013.16.3.254 (cit. on p. 5).

[19]   Y. Jiménez-Gaona, M. J. Rodríguez-Álvarez, and V. Lakshminarayanan. "Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review". In: *Applied Sciences* 10.22 (2020). doi: 10.3390/app10228298 (cit. on pp. 5, 10, 19).

[20]   M. K. Hasan and S. R. Ara. *Detection and classification of breast lesions using ultrasound-based imaging modalities*. Vol. 3. Elsevier, 2019, pp. 331–348. doi: 10.1016/B978-0-12-801238-3.64169-7 (cit. on pp. 5, 12).

[21]   Y. M. Lei et al. "Artificial Intelligence in Medical Imaging of the Breast". In: *Frontiers in Oncology* 11.July (2021), pp. 1–10. doi: 10.3389/fonc.2021.600557 (cit. on pp. 5, 10, 16, 17, 19, 21).

[22] A. Ciritsis et al. "Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making". In: *European Radiology* 29.10 (2019), pp. 5458–5468. doi: 10.1007/s00330-019-06118-7 (cit. on p. 6).

[23] N. I. Yassin et al. "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review". In: *Computer Methods and Programs in Biomedicine* 156 (2018), pp. 25–45. doi: 10.1016/j.cmpb.2017.12.012 (cit. on pp. 6, 10, 17).

[24] S. V. Sree. "Breast imaging: A survey". In: *World Journal of Clinical Oncology* 2.4 (2011), p. 171. doi: 10.5306/wjco.v2.i4.171 (cit. on p. 6).

[25] H. Li et al. "Benign and malignant classification of mammogram images based on deep learning". In: *Biomedical Signal Processing and Control* 51 (2019), pp. 347–354. doi: 10.1016/j.bspc.2019.02.017. url: https://doi.org/10.1016/j.bspc.2019.02.017 (cit. on p. 6).

[26] R. M. Mann, C. K. Kuhl, and L. Moy. "Contrast-enhanced MRI for breast cancer screening". In: *Journal of Magnetic Resonance Imaging* 50.2 (2019), pp. 377–390. doi: 10.1002/jmri.26654 (cit. on p. 6).

[27] Y. S. Sun et al. "Risk factors and preventions of breast cancer". In: *International Journal of Biological Sciences* 13.11 (2017), pp. 1387–1397. doi: 10.7150/ijbs.21635 (cit. on p. 6).

[28] M. Welleweerd et al. "Robot-assisted ultrasound-guided biopsy on MR-detected breast lesions". In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct. 2020, pp. 2965–2971. doi: 10.1109/IROS45743.2020.9341695 (cit. on pp. 6, 11, 12).

[29] H. Cai et al. "Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: A machine learning approach". In: *PLoS ONE* 9.1 (2014). doi: 10.1371/journal.pone.0087387 (cit. on pp. 6, 7).

[30] B. Behboodi et al. "Deep classification of breast cancer in ultrasound images: more classes, better results with multi-task learning". In: (2021), p. 3. doi: 10.1117/12.2581930 (cit. on pp. 6, 21, 25, 46, 54).

[31] D. Thigpen, A. Kappler, and R. Brem. "The role of ultrasound in screening dense breasts - A review of the literature and practical solutions for implementation". In: *Diagnostics* 8.1 (2018), pp. 1–14. doi: 10.3390/diagnostics8010020 (cit. on p. 6).

[32] C. Xue et al. "Global guidance network for breast lesion segmentation in ultrasound images". In: *Medical Image Analysis* 70 (2021), p. 101989. doi: 10.1016/j.media.2021.101989 (cit. on pp. 6, 20, 24, 31–33, 37, 42, 73).

[33] Kriti, J. Virmani, and R. Agarwal. "Deep feature extraction and classification of breast ultrasound images". In: *Multimedia Tools and Applications* 79.37-38 (Oct. 2020), pp. 27257–27292. doi: 10.1007/s11042-020-09337-z (cit. on pp. 7, 8).

[34] E. Uzunhisarcikli and V. Goreke. "A novel classifier model for mass classification using BI-RADS category in ultrasound images based on Type-2 fuzzy inference system". In: *Sadhana - Academy Proceedings in Engineering Sciences* 43.9 (2018), pp. 1–12. doi: `10.1007/s12046-018-09 15-x` (cit. on p. 7).

[35] C. J. D'Orsi. *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. American College of Radiology, 2013. isbn: 9781559030168 (cit. on pp. 7, 8).

[36] S. J. K. Hsieh et al. *Manual do Residente de Imaginologia Mamária do InRad*. 1st ed. Santana de Parnaíba: Manole, 2022, p. 296 (cit. on p. 9).

[37] B. Shareef, M. Xian, and A. Vakanski. "Stan: Small Tumor-Aware Network for Breast Ultrasound Image Segmentation". In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. Vol. 2020-April. IEEE, Apr. 2020, pp. 1–5. doi: `10.1109/ISBI45749.2020.9098691` (cit. on pp. 10, 19, 24).

[38] G.-G. Wu et al. "Artificial intelligence in breast ultrasound." In: *World journal of radiology* 11.2 (Feb. 2019), pp. 19–26. doi: `10.4329/wjr.v11.i2.19` (cit. on pp. 10, 22, 60).

[39] Y. Wu et al. "BGM-Net: Boundary-Guided Multiscale Network for Breast Lesion Segmentation in Ultrasound". In: *Frontiers in Molecular Biosciences* 8.July (2021), pp. 1–8. doi: `10.3389/fmolb.2021.698334` (cit. on p. 10).

[40] I. C. Education. *What are neural networks?* url: `https://www.ibm.com/uk-en/cloud/learn/neural-networks` (cit. on pp. 10, 11).

[41] L. Shukla. *Designing your neural networks*. Sept. 2019. url: `https://towardsdatascience.com/designing-your-neural-networks-a5e4617027ed` (cit. on p. 10).

[42] A. Mathew, P. Amudha, and S. Sivakumari. "Deep learning techniques: an overview". In: *Advances in Intelligent Systems and Computing* 1141 (Jan. 2021), pp. 599–608. doi: `10.1007/978-98 1-15-3383-9_54` (cit. on p. 11).

[43] G. Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical Image Analysis* 42.December 2012 (2017), pp. 60–88. doi: `10.1016/j.media.2017.07.005` (cit. on pp. 11, 17).

[44] U. Bick et al. "Image-guided breast biopsy and localisation: recommendations for information to women and referring physicians by the European Society of Breast Imaging". In: *Insights into Imaging* 11.1 (2020). doi: `10.1186/s13244-019-0803-x` (cit. on pp. 11, 12).

[45] M. Nolan-Pleckham. *What Is a Breast Biopsy?* Ed. by D. Paul. url: `https://www.verywellhealth.com/open-surgical-breast-biopsy-429949` (cit. on p. 11).

[46] M. S. Newell and M. C. Mahoney. "Ultrasound-guided percutaneous breast biopsy". In: *Techniques in vascular and interventional radiology* 17.1 (Mar. 2014), pp. 23–31. doi: `10.1053/J.TVIR.2 013.12.005` (cit. on p. 12).

[47] Y. Hu et al. "Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model". In: *Medical Physics* 46.1 (2019), pp. 215–228. doi: 10.1002/mp.13268 (cit. on pp. 12, 31, 45).

[48] V. K. Singh et al. "An Efficient Solution for Breast Tumor Segmentation and Classification in Ultrasound Images Using Deep Adversarial Learning". In: (2019), pp. 1–9 (cit. on pp. 12, 60).

[49] X. Liu et al. "A review of deep-learning-based medical image segmentation methods". In: *Sustainability (Switzerland)* 13.3 (2021), pp. 1–29. doi: 10.3390/su13031224 (cit. on p. 12).

[50] T. Fernando et al. "Deep Learning for Medical Anomaly Detection A Survey". In: *ACM Computing Surveys* 54.7 (2022), pp. 1–28. doi: 10.1145/3464423 (cit. on pp. 16, 17).

[51] H. Zenati et al. "Adversarially Learned Anomaly Detection". In: *2018 IEEE International Conference on Data Mining (ICDM)*. Vol. 2018-Novem. IEEE, Nov. 2018, pp. 727–736. doi: 10.1109/ICDM.2018.00088 (cit. on p. 16).

[52] M. Awad and R. Khanna. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Apress, Jan. 2015, pp. 1–248. doi: 10.1007/978-1-4302-5990-9 (cit. on p. 16).

[53] L. Cai, J. Gao, and D. Zhao. "A review of the application of deep learning in medical image classification and segmentation". In: *Annals of Translational Medicine* 8.11 (2020), pp. 713–713. doi: 10.21037/atm.2020.02.44 (cit. on p. 17).

[54] Kriti, J. Virmani, and R. Agarwal. *A Review of Segmentation Algorithms Applied to B-Mode Breast Ultrasound Images: A Characterization Approach*. Vol. 28. 4. Springer Netherlands, 2021, pp. 2567–2606. doi: 10.1007/s11831-020-09469-3 (cit. on p. 17).

[55] Z. Zhang et al. "Ultrasonic Diagnosis of Breast Nodules Using Modified Faster R-CNN". In: *Ultrasonic Imaging* 41.6 (2019), pp. 353–367. doi: 10.1177/0161734619882683 (cit. on p. 17).

[56] M. I. Razzak, S. Naz, and A. Zaib. "Deep Learning for Medical Image Processing: Overview, Challenges and the Future". In: *Springer* (2018), pp. 323–350. doi: https://doi.org/10.1007/978-3-319-65981-7_12 (cit. on p. 17).

[57] L. Wang et al. "Ai auxiliary labeling and classification of breast ultrasound images". In: *Journal of Image and Graphics(United Kingdom)* 9.2 (2021), pp. 45–49. doi: 10.18178/joig.9.2.45-49 (cit. on pp. 17, 21, 25, 73).

[58] M. H. Hesamian et al. "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges". In: *Journal of Digital Imaging* 32.4 (2019), pp. 582–596. doi: 10.1007/s10278-019-00227-x (cit. on p. 17).

[59] S. K. Zhou, H. Greenspan, and D. Shen. "Deep Learning for Medical Image Analysis". In: *Deep Learning for Medical Image Analysis* (2017), pp. 1–433 (cit. on p. 17).

[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems* 25 (2012) (cit. on p. 17).

[61] C. Szegedy et al. "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 07-12-June. IEEE, June 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594 (cit. on pp. 17, 21, 45, 47, 48).

[62] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, Sept. 2015 (cit. on pp. 17, 21).

[63] C. Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2016-Decem. IEEE Computer Society, Oct. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308 (cit. on pp. 17, 21, 46–49).

[64] K. He et al. "Spatial pyramid pooling in deep convolutional networks for visual recognition". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8691 LNCS. PART 3. Springer Verlag, June 2014, pp. 346–361. doi: 10.1007/978-3-319-10578-9_23 (cit. on p. 17).

[65] K. He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2016-Decem. IEEE Computer Society, Dec. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90 (cit. on pp. 17, 18, 21, 23, 46–48, 60, 62).

[66] G. Huang et al. "Densely connected convolutional networks". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Vol. 2017-Janua. Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243 (cit. on pp. 17, 18, 21, 23, 46–50).

[67] M. Sandler et al. "MobileNetV2: Inverted Residuals and Linear Bottlenecks". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2018), pp. 4510–4520. doi: 10.1109/CVPR.2018.00474 (cit. on pp. 17, 46–50).

[68] M. Tan and Q. V. Le. "EfficientNet: Rethinking model scaling for convolutional neural networks". In: *36th International Conference on Machine Learning, ICML 2019* (2019), pp. 10691–10700 (cit. on pp. 17, 46–48, 50).

[69] J. Long, E. Shelhamer, and T. Darrell. "Fully Convolutional Networks for Semantic Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4 (Nov. 2017), pp. 640–651. doi: 10.1109/TPAMI.2016.2572683 (cit. on pp. 18, 60).

[70] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab et al. Springer International Publishing, 2015, pp. 234–241 (cit. on pp. 18, 23, 32–34, 61).

[71] O. Oktay et al. "Attention U-Net: Learning Where to Look for the Pancreas". In: Midl (2018) (cit. on p. 18).

[72] Z. Zhou et al. "Unet++: A nested u-net architecture for medical image segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11045 LNCS (2018), pp. 3–11. doi: 10.1007/978-3-030-00889-5 _1 (cit. on p. 18).

[73] F. Isensee et al. "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation". In: *ArXiv* (2018). doi: 10.1007/978-3-658-25326-4_7. eprint: 1809.10486 (cit. on pp. 18, 32–34).

[74] A. Myronenko. "3D MRI brain tumor segmentation using autoencoder regularization". In: *Brain-Les@MICCAI*. 2018 (cit. on pp. 18, 32–35, 62).

[75] A. Dosovitskiy et al. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020) (cit. on pp. 18, 35).

[76] A. Hatamizadeh et al. "UNETR: Transformers for 3D Medical Image Segmentation". In: (2021) (cit. on pp. 18, 32, 33, 35, 36, 42).

[77] Q. Kang et al. "Thyroid nodule segmentation and classification in ultrasound images through intra- and inter-task consistent learning". In: *Medical Image Analysis* 79 (2022), p. 102443. doi: 10.1016/j.media.2022.102443 (cit. on pp. 18, 61, 70, 75).

[78] T.-L.-T. Le et al. "Multitask Classification and Segmentation for Cancer Diagnosis in Mammography". In: 2017 (2019), pp. 1–4 (cit. on pp. 18, 19, 60, 61).

[79] Y. Zhou et al. "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images". In: *Medical Image Analysis* 70 (2021), p. 101918. doi: 10.1016/j.media.2020.101918 (cit. on pp. 19, 60, 61).

[80] K. Wang, S. Liang, and Y. Zhang. "Residual Feedback Network for Breast Lesion Segmentation in Ultrasound Image". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12901 LNCS (2021), pp. 471–481. doi: 10.1007/978-3-030-87193-2_45 (cit. on pp. 19, 20, 24, 32, 33, 36).

[81] Z. Zhuang et al. "An RDAU-NET model for lesion segmentation in breast ultrasound images". In: *PLoS ONE* 14.8 (2019), pp. 1–23. doi: 10.1371/journal.pone.0221535 (cit. on pp. 19, 24, 28, 29, 40).

[82]   A. Negi et al. "RDA-UNET-WGAN: An Accurate Breast Ultrasound Lesion Segmentation Using Wasserstein Generative Adversarial Networks". In: *Arabian Journal for Science and Engineering* 45.8 (2020), pp. 6399–6410. doi: 10.1007/s13369-020-04480-z (cit. on pp. 19, 24).

[83]   B. Shareef et al. "ESTAN: Enhanced Small Tumor-Aware Network for Breast Ultrasound Image Segmentation". In: *IEEE Transactions on Medical Imaging* (2020) (cit. on pp. 20, 24).

[84]   V. K. Singh et al. "Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework". In: *Expert Systems with Applications* 162.February (2020), p. 113870. doi: 10.1016/j.eswa.2020.113870 (cit. on p. 20).

[85]   L. Zhu et al. *A Second-Order Subregion Pooling Network for Breast Lesion Segmentation in Ultrasound.* Vol. 12266 LNCS. 10. Springer International Publishing, 2020, pp. 160–170. doi: 10.1007/978-3-030-59725-2_16 (cit. on pp. 20, 24).

[86]   M. Xu et al. "Mssa-Net: Multi-Scale Self-Attention Network For Breast Ultrasound Image Segmentation". In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. Vol. 2021-April. 3. IEEE, Apr. 2021, pp. 827–831. doi: 10.1109/ISBI48211.2021.9433899 (cit. on pp. 20, 24, 39, 73).

[87]   A. Iqbal and M. Sharif. "MDA-Net: Multiscale dual attention-based network for breast lesion segmentation using ultrasound images". In: *Journal of King Saud University - Computer and Information Sciences* November (2021). doi: 10.1016/j.jksuci.2021.10.002 (cit. on pp. 20, 24, 32, 33, 37, 73).

[88]   N. S. Punn and S. Agarwal. "RCA-IUnet: a residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging". In: *Machine Vision and Applications* 33.2 (Mar. 2022), p. 27. doi: 10.1007/s00138-022-01280-3 (cit. on pp. 20, 24, 73).

[89]   S. Han et al. "A deep learning framework for supporting the classification of breast lesions in ultrasound images". In: *Physics in Medicine & Biology* 62.19 (Sept. 2017), pp. 7714–7728. doi: 10.1088/1361-6560/aa82ec (cit. on pp. 21, 25, 45, 46, 54).

[90]   T. Xiao et al. "Comparison of Transferred Deep Neural Networks in Ultrasonic Breast Masses Discrimination". In: *BioMed Research International* 2018 (2018). doi: 10.1155/2018/4605191 (cit. on pp. 21, 25, 46, 54).

[91]   F. Chollet. "Xception: Deep learning with depthwise separable convolutions". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Vol. 2017-Janua. Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195 (cit. on p. 21).

[92]   M. Byra et al. "Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion". In: *Medical Physics* 46.2 (2019), pp. 746–755. doi: 10.1002/mp.13361 (cit. on pp. 21, 25).

[93]   S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *32nd International Conference on Machine Learning, ICML 2015*. Vol. 1. International Machine Learning Society (IMLS), Feb. 2015, pp. 448–456 (cit. on p. 21).

[94]   T. Fujioka et al. "Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network". In: *Japanese Journal of Radiology* 37.6 (2019), pp. 466–472. doi: 10.1007/s11604-019-00831-5 (cit. on pp. 21, 25, 45).

[95]   W. K. Moon et al. "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks". In: *Computer Methods and Programs in Biomedicine* 190.105361 (2020). doi: 10.1016/j.cmpb.2020.105361 (cit. on pp. 21, 22, 25, 46, 51, 54, 73).

[96]   S. Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.6 (June 2017), pp. 1137–1149. doi: 10.1109/TPAMI.2016.2577031 (cit. on pp. 21, 23).

[97]   E. Y. Kalafi et al. "Classification of breast cancer lesions in ultrasound images by using attention layer and loss ensemble in deep convolutional neural networks". In: *Diagnostics* 11.10 (2021), pp. 1–14. doi: 10.3390/diagnostics11101859 (cit. on pp. 21, 22, 25, 73).

[98]   W. Cui et al. "FMRNet: A fused network of multiple tumoral regions for breast tumor classification with ultrasound images". In: *Medical Physics* (2021). doi: 10.1002/mp.15341 (cit. on pp. 22, 25, 45, 73).

[99]   H. Ma et al. "Fus2Net: a novel Convolutional Neural Network for classification of benign and malignant breast tumor in ultrasound images". In: *BioMedical Engineering Online* 20.1 (2021), pp. 1–15. doi: 10.1186/s12938-021-00950-z (cit. on pp. 22, 25, 38, 45, 73).

[100] G. Zhang et al. "LRSCnet: Local Reference Semantic Code learning for breast tumor classification in ultrasound images". In: *BioMedical Engineering OnLine* 20.1 (2021), pp. 1–14. doi: 10.1186/s12938-021-00968-3 (cit. on pp. 22, 25, 45, 73).

[101] K. He et al. "Mask R-CNN". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.2 (Mar. 2020), pp. 386–397. doi: 10.1109/TPAMI.2018.2844175 (cit. on p. 22).

[102] G. Zhang et al. "SHA-MTL: soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification". In: *International Journal of Computer Assisted Radiology and Surgery* 16.10 (2021), pp. 1719–1725. doi: 10.1007/s11548-021-02445-7 (cit. on pp. 23, 25).

[103] Y. Luo, Q. Huang, and X. Li. "Segmentation information with attention integration for classification of breast tumor in ultrasound image". In: *Pattern Recognition* (2021), p. 108427. doi: 10.1016/j.patcog.2021.108427 (cit. on pp. 23, 25).

[104] Z. Cao et al. "An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures". In: *BMC Medical Imaging* 19.1 (2019), pp. 1–9. doi: 10.1186/s12880-019-0349-x (cit. on p. 26).

[105] W. Al-Dhabyani et al. "Dataset of breast ultrasound images". In: *Data in Brief* 28 (2020), p. 104863. doi: 10.1016/j.dib.2019.104863 (cit. on p. 27).

[106] M. H. Yap et al. "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks". In: *IEEE Journal of Biomedical and Health Informatics* 22.4 (2018), pp. 1218–1226. doi: 10.1109/JBHI.2017.2731873 (cit. on pp. 27, 28).

[107] *MATLAB - MathWorks - MATLAB & Simulink*. url: https://www.mathworks.com/products/matlab.html (cit. on p. 30).

[108] L. Liu et al. "Automated breast tumor detection and segmentation with a novel computational framework of whole ultrasound images". In: *Medical & Biological Engineering & Computing* 56.2 (2018), pp. 183–199. doi: 10.1007/s11517-017-1770-3 (cit. on p. 31).

[109] *Project MONAI*. url: https://docs.monai.io/en/stable/index.html# (cit. on p. 32).

[110] R. A. Zeineldin et al. "Ensemble CNN Networks for GBM Tumors Segmentation Using Multiparametric MRI". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 12962 LNCS. 2022, pp. 473–483. isbn: 9783031089985. doi: 10.1007/978-3-031-08999-2_41. url: https://link.springer.com/10.1007/978-3-031-08999-2_41 (cit. on p. 35).

[111] S. Jadon. "A survey of loss functions for semantic segmentation". In: *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, Oct. 2020, pp. 1–7. isbn: 978-1-7281-9468-4. doi: 10.1109/CIBCB48159.2020.9277638. eprint: 2006.14822 (cit. on p. 38).

[112] J. Ma et al. "Loss odyssey in medical image segmentation". In: *Medical Image Analysis* 71 (July 2021), p. 102035. doi: 10.1016/j.media.2021.102035 (cit. on p. 38).

[113] F. Milletari, N. Navab, and S.-A. Ahmadi. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, Oct. 2016, pp. 565–571. doi: 10.1109/3DV.2016.79 (cit. on pp. 38, 61).

[114] F. Isensee et al. "Automated Design of Deep Learning Methods for Biomedical Image Segmentation". In: (Apr. 2019), pp. 1–55. url: http://arxiv.org/abs/1904.08128 (cit. on p. 38).

[115] *PyTorch*. url: https://pytorch.org/ (cit. on p. 39).

[116] A. A. Taha and A. Hanbury. "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool". In: *BMC Medical Imaging* 15.1 (2015). doi: 10.1186/s12880-015-0068-x (cit. on pp. 39, 40).

[117] Y. Gu et al. "Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study". In: *Insights into Imaging* 13.1 (Dec. 2022), p. 124. issn: 1869-4101. doi: 10.1186/s13244-022-01259-8 (cit. on pp. 45, 57).

[118] *Inception-v3 Explained - Papers With Code*. url: https://paperswithcode.com/method/inception-v3 (cit. on p. 48).

[119] S. Mukherjee. *The Annotated ResNet-50*. Aug. 2022. url: https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758 (cit. on p. 49).

[120] T. Ahmed and N. H. N. Sabab. "Classification and Understanding of Cloud Structures via Satellite Images with EfficientUNet". In: *SN Computer Science* 3.1 (2022). issn: 2662-995X. doi: 10.1007/s42979-021-00981-2. arXiv: 2009.12931 (cit. on p. 50).

[121] T. Ahmed and N. Sabab. "Classification and understanding of cloud structures via satellite images with EfficientUNet". In: (Sept. 2020). doi: 10.1002/essoar.10507423.1 (cit. on p. 51).

[122] S. Gokhale. "Ultrasound characterization of breast masses". In: *Indian Journal of Radiology and Imaging* 19.03 (July 2009), pp. 242–247. issn: 0971-3026. doi: 10.4103/0971-3026.54878. url: http://www.thieme-connect.de/DOI/DOI?10.4103/0971-3026.54878 (cit. on pp. 52, 56).

[123] A. Tharwat. "Classification assessment methods". In: *Applied Computing and Informatics* 17.1 (2018), pp. 168–192. doi: 10.1016/j.aci.2018.08.003 (cit. on p. 53).

[124] Y. W. Chang et al. "A novel computer-aided-diagnosis system for breast ultrasound images based on BI-RADS categories". In: *Applied Sciences (Switzerland)* 10.5 (2020), pp. 1–20. doi: 10.3390/app10051830 (cit. on p. 53).

[125] M. Grandini, E. Bagli, and G. Visani. "Metrics for Multi-Class Classification: an Overview". In: (2020), pp. 1–17. url: http://arxiv.org/abs/2008.05756 (cit. on p. 53).

[126] M. Norton and S. Uryasev. "Maximization of AUC and Buffered AUC in binary classification". In: *Mathematical Programming* 174.1-2 (2019), pp. 575–612. doi: 10.1007/s10107-018-1312-2. url: https://doi.org/10.1007/s10107-018-1312-2 (cit. on p. 53).

[127] S. Ruder. "An Overview of Multi-Task Learning in Deep Neural Networks". In: (May 2017). url: http://arxiv.org/abs/1706.05098 (cit. on p. 60).

[128] S. Gang et al. "Character Recognition of Components Mounted on Printed Circuit Board Using Deep Learning". In: *Sensors* 21.9 (Apr. 2021), p. 2921. doi: 10.3390/s21092921 (cit. on p. 63).